```
OM of: US-09-303-518D-649 to: A_Geneseq_032802:*   out_format : pfs

Date: Jun 30, 2002  6:48 AM

About: Results were produced by the GenCore software, version 4.5,
       Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
 -MODEL=frame+_n2p.model -DEV=xlh
 -Q=/cgn2_1/USPTO_spool/US09303518/runat_28062002_142712_4291/app_query.fasta_1.23501
 -DB=A_Geneseq_032802 -QFMT=fastan -SUFFIX=rag -GAPOP=12.000
 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
 -DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
 -TRANS=human40.cdi -LIST=100 -DOCALIGN=200 -THR_SCORE=pct
 -THR_MAX=100 -THR_MIN=0 -ALIGN=45 -MODE=LOCAL -OUTFMT=pfs
 -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
 -USER=US09303518_@CGN1_1_0 -NCPU=6 -ICPU=3 -LONGLOG
 -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-303-518D-649
Query length: 4374
Database: A_Geneseq_032802:*
Database sequences: 747574
Database length: 111073796
Search time (sec): 627.340000


score_list:
Sequence                 Strd Orig    ZScore    EScore Len  ! Documentation  ..
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:AAR07304 + 1608.50 2079.18   7.4e-109  15
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY75564 + 1547.00 1999.29   2.2e-104  14
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABB52745 + 1071.00 1375.04   1.4e-69   13
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABB52592 + 1053.00 1351.37   2.9e-68   13
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:AAW27705 +  715.50  924.55   7.3e-44   32
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:AAP80136 +  420.00  527.48   4.1e-22   74
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAG98842 +  364.50  446.29   6.5e-18   15
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01830 +  307.00  373.61   9.4e-14   12
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01828 +  307.00  373.56   9.4e-14   12
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY81609 +  298.00  361.66   4.3e-13   12
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:AAR26503 +  297.50  364.41   4.1e-13   91
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:AAR14320 +  295.50  361.79   5.7e-13   91
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:AAR25578 +  295.00  361.00   6.3e-13   92
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:AAR41732 +  289.50  348.15   2.0e-12   15
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:AAW30292 +  286.50  343.70   3.3e-12   16
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:AAR63506 +  285.00  342.63   4.1e-12   14
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:AAR41724 +  283.50  340.66   5.3e-12   14
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:AAR41728 +  283.00  340.00   5.8e-12   14
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01848 +  283.00  340.00   5.8e-12   14
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:AAW30291 +  281.50  337.16   7.7e-12   15
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:AAR41731 +  281.00  338.48   7.7e-12   13
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01827 +  280.50  341.41   7.3e-12   96
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01826 +  280.50  341.34   7.4e-12   97
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:AAW30294 +  278.50  334.10   1.2e-11   14
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:AAR63505 +  277.50  332.35   1.5e-11   15
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01846 +  277.50  332.35   1.5e-11   15
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01841 +  274.00  332.49   2.2e-11   10
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01840 +  274.00  332.42   2.2e-11   10
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:AAR41723 +  272.50  325.79   3.4e-11   15
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:AAW30293 +  269.50  321.85   5.7e-11   15
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01837 +  268.00  323.87   6.3e-11   10
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01836 +  267.50  323.15   6.9e-11   10
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:AAR41725 +  267.50  319.22   8.0e-11   15
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01825 +  265.00  318.50   1.1e-10   12
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01824 +  265.00  318.44   1.1e-10   12
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01835 +  264.50  319.05   1.1e-10   10
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01834 +  264.50  318.99   1.1e-10   11
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAU34143 +  264.50  311.88   1.5e-10   20
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAU37017 +  264.50  300.51   2.3e-10   57
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAU37403 +  262.50  296.99   3.4e-10   62
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAU37120 +  261.50  306.65   2.6e-10
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01843 +  261.00  315.56   2.0e-10
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01842 +  261.00  315.49   2.0e-10
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01833 +  259.50  313.44   2.6e-10
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01832 +  259.50  313.38   2.6e-10
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG20365 +  258.00  310.49   3.4e-10
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY75097 +  257.50  300.62   5.3e-10
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01847 +  254.50  305.93   6.1e-10
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY75096 +  254.50  298.96   8.0e-10
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY75098 +  254.00  295.65   9.7e-10
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01849 +  251.50  302.61   9.9e-10
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG30355 +  249.50  295.10   1.7e-09
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABG24938 +  243.50  291.54   3.9e-09
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAU37018 +  243.50  279.70   6.2e-09
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:AAY27230 +  243.00  284.25   5.5e-09
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB59826 +  242.00  285.20   6.0e-09
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01839 +  241.50  290.36   5.2e-09
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01838 +  241.50  290.30   5.2e-09
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:AAW65088 +  241.00  283.91   7.0e-09
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAG98256 +  240.50  285.43   7.0e-09
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB23860 +  240.50  278.77   9.1e-09
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB59813 +  240.00  287.72   6.8e-09
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABB52597 +  239.00  289.42   7.2e-09
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01844 +  239.00  284.68   8.6e-09
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01845 +  238.00  283.44   1.0e-08
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:AAW04505 +  238.00  277.61   1.3e-08
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB69133 +  238.00  277.61   1.3e-08
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB69137 +  238.00  277.61   1.3e-08
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB69134 +  238.00  277.31   1.3e-08
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:AAW42634 +  234.00  272.43   2.5e-08
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAU27577 +  232.00  273.30   3.1e-08
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAE10024 +  232.00  273.30   3.1e-08
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB69136 +  231.50  267.42   4.1e-08
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:AAW20725 +  230.00  270.94   4.2e-08
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAU34338 +  230.00  264.87   5.3e-08
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:AAR99393 +  229.50  264.60   5.7e-08
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:AAW98391 +  229.00  265.90   5.8e-08
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:AAW56322 +  229.00  265.54   5.9e-08
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:AAY11001 +  228.50  269.21   5.4e-08
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAU34412 +  227.00  268.68   6.6e-08
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAU37520 +  227.00  268.19   6.7e-08
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB46351 +  227.00  258.99   9.5e-08
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAB69135 +  226.50  262.19   9.0e-08
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:AAW56319 +  224.00  258.96   1.4e-07
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:AAR05041 +  224.00  252.51   1.7e-07
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABB69806 +  223.00  259.09   1.5e-07
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAE00701 +  223.00  257.22   1.6e-07
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:AAW98828 +  222.50  253.12   2.0e-07
/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:AAW71556 +  222.50  253.12   2.0e-07

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:AAR07304

seq_documentation_block:
ID   AAR07304 standard; protein; 1541 AA.
XX
AC   AAR07304;
XX
DT   31-JAN-1991  (first entry)
XX
DE   IgA1 protease.
XX
KW   IgA1; vaccine; meningitis; gonnorhoea; allergies.
XX
OS   Haemophilus influenzae.
XX
PN   WO9011367-A.
XX
PD   04-OCT-1990.
XX
PF   16-MAR-1990;    90WO-DK00073.
XX
PR   17-MAR-1989;    89DK-0001308.
XX
```

```
PA   (KILI/) KILIAN M.
XX
PI   Kilian M,  Poulsen K;
XX
DR   WPI; 1990-320267/42.
DR   N-PSDB; AAQ06164.
XX
PT   Immunoglobulin A1 protease prodn. - by cloning from
PT   microorganisms for immunisation against immunoglobulin A1
PT   protease producing bacteria
XX
PS   Disclosure; fig 3; 44pp; English.
XX
CC   This immunoglobulin (Ig)A1 protease is produced by recombinant DNA
CC   methods.  It is useful in a vaccine for e.g. meningococcal meningitis,
CC   gonorrhoea or allergic diseases.  It specifically cleaves the heavy
CC   chain of human IgA1 in the hinge region.
XX
SQ   Sequence   1541 AA;

alignment_scores:
           Quality: 1608.50         Length:   1717
             Ratio:  1.733            Gaps:     51
  Percent Similarity:  54.048   Percent Identity: 27.257

alignment_block:
  US-09-303-518D-649 x AAR07304   ..

  Align seg 1/1  to: AAR07304  from: 1  to: 1541

     64 CGCTTCTCGCCTGCTTACTTAGCCATATGCCTGTCGTTCGGCATTCTTCC 113
         :::|||:::       ::::::|||:::    ::::::::::::::    ||
      5 LysPheLysLeuAsnPheIleAlaLeuThrValAlaTyrAlaLeuThrPr 21

    114 CCAAGCCTGGGCGGGACACACTTATTTCGGCATCAACTACCAATACTATC 163
         |    :::    |||:::    :::         ::::::||||||   :::|
     21 oTyrThrGluAlaAlaLeuValArgAspAspValAspTyrGlnIlePheA 38

    164 GCGACTTTGCCGAAAATAAAGGCAAGTTTGCAGTCGGGGCGAAAGATATT 213
         ||||||||||||||||||||||||||||||:::|||||||||    ::::::
     38 rgAspPheAlaGluAsnLysGlyLysPheSerValGlyAlaThrAsnVal 54

    214 GAGGTTTACAACAAAAAAGGGGAGTTGGTCGGCAAATCAATGACAAAAGC 263
          |||     :::|||:::::::::     :::|||   ::::::   :::::
     55 LeuValLysAspLysAsnAsnLysAspLeuGlyThrAlaLeuProAsnGl 71

    264 C...CCGATGATTGATTTTTCTGTGGTGTCG...CGTAACGGCGTGGCGG 307
        :     |||||||||||||||||||||||||:::        :::    :::|||:
     71 yIleProMetIleAspPheSerValValAspValAspLysArgIleAlaT 88

    308 CATTGGTGGGCGATCAATATATTGTGAGCGTGGCACAT........... 345
         ::|||:::::      ||||||:::|||:::|||     |||
     88 hrLeuIleAsnProGlnTyrValValGlyValLysHisValSerAsnGly 104

    346 .........................AACGGCGGCTATAACAACGTTGA 368
                                  |||||||:::    ||||||     ::
    105 ValSerGluLeuHisPheGlyAsnLeuAsnGlyAsnMetAsnAsnGlyAs 121

    369 TTTTGGTGCGGAAGGAAGAAATCCCGATCAACATCGTTTTACTTATAAAA 418
         :      |||:::     |||:::     ::::::::::::        |||
    121 nAlaLysAlaHis...ArgAspValSerSerGluGluAsnArgTyrPheS 137

    419 TTGTGAAACGGAATAATTATAAAGCAGGGACTAAAGGC........... 456
           |||::::::|||:::|||     :::         :::|||
    137 erValGluLysAsnGluTyrProThrLysLeuAsnGlyLysThrValThr 153

    457 ..............CATCCTTATGGCGGCGATTATCATATGCCGCGTTT 491
                      :::                  |||||||:::||||||||||||
    154 ThrGluAspGlnThrGlnLysArgArgGluAspTyrTyrMetProArgLe 170

    492 GCATAAATTTGTCACAGATGCAGAACCTGTTGAA...ATGACCAGTTATA 538
         |    ||||||||||||:::::     |||:::|||       |||:::
    170 uAspLysPheValThrGluValAlaProIleGluAlaSerThrAlaSerS 187

    539 TGGATGGGCGGAAATATATCGATCAAAATAATTACCCTGACCGTGTTCGT 588
         |||:::       |||    |||||||||:::||||||         ||||||
    187 erAspAlaGlyThrTyrAsnAspGlnAsnLysTyrProAlaPheValArg 203

    589 ATTGGGGCAGGCAGGCAATATTGGCGATCTGATGAAGATGAG....... 630
         :::|||:::|||    |||:::         :::         |||:::
    204 LeuGlySerGlySerGlnPheIleTyrLysLysGlyAspAsnTyrSerLe 220

    631 ....CCCAATAACCGCGAAAGTTCATATCAT...........ATTGCAA 664
                |||||||:::|||    :::    :::              ::::::
    220 uIleLeuAsnAsnHisGluValGlyGlyAsnAsnLeuLysLeuValGlyA 237

    665 GTGCGTATTCTTGGCTCGTTGGTGGCAATACCTTTGCACAAAATGGATCA 714
         ::||||||:::::     ::::::|||:::     :::         |||    :::
    237 spAlaTyrThrTyrGlyIleAlaGlyThrProTyrLysValAsnHisGlu 253

    715 GGTGGTGGCACAGTCAACTTAGGTAGTGAAAAAATTAAACATAGC...CC 761
         ::::::|||     :::::::::|||::::::|||     :::||||||     ||
    254 AsnAsnGlyLeuIleGlyPheGlyAsnSerLysGluGluHisSerAspPr 270

    762 ATATGGTTTTTTACCAACAGGA..............GGCTCATTTGGCG 796
         |    |||:::|||                         :::    :::||||
    270 oLysGlyIleLeuSerGlnAspProLeuThrAsnTyrAlaValLeuGlyA 287

    797 ACAGTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCAAAAGTGGTTA 846
         ||||||||||||||:::|||:::||||||     :::|||     ||||||||
    287 spSerGlySerProLeuPheValTyrAspArgGluLysGlyLysTrpLeu 303

    847 ATTAATGGGGTATTGCAAACGGGCAACCCCTATATAGGAAAAAGCAATGG 896
         :::    |||     :::               |||     ||||||
    304 PheLeuGlySerTyrAspPheTrpAlaGlyTyrAsnLysLysSer..... 318

    897 CTTCCAGCTGGTTCGTAAAGATTGG...TTCTATGATGAAATCTTTGCTG 943
                     ::::::|||     :::|||     :::     |||:::
    319 .............TrpGlnGluTrpAsnIleTyrLysSerGlnPheThrL 331

    944 GAGATACCCATTCAGTATTCTACGAACCACGTCAAAATGGGAAATACTCT 993
          |||:::
    331 ysAspVal.......................................... 333

    994 TTTAACGACGATAATAATGGCACAGGAAAAATCAATGCCAAACATGAACA 1043
         :::|||     |||:::     |||:::          |||::::::|||     :::::
    334 LeuAsnLysAspSerAlaGlySer...LeuIleGlySerLysThrAspTy 349

   1044 CAATTCTCTGCCTAATAGATTAAAAACACGAACCGTT...........C 1081
         ::::            |||         ||||||     |||:::              :
    349 rSerTrpSerSerAsnGly...LysThrSerThrIleThrGlyGlyGluL 365

   1082 AATTGTTTAATGTTTCTTTATCCGAGACAGCAAGAGAACCTGTTTATCAT 1131
         ::     :::||||||:::|||::::::            :::|||
    365 ysSerLeuAsnValAspLeuAlaAspGlyLysAspLysPro......... 378

   1132 GCTGCAGGTGGTGTCAACAGTTATCGACCCAGACTGAATAATGGAGAAAA 1181
                                                 |||:::|||:::::
    379 ..................................AsnHisGlyLysSe 383

   1182 TATTTCCTTTATTGACGAAGGAAAAGGCGAATTGATACTTACCAGCAACA 1231
         :::::::|||          |||||||:::|||     |||     |||:::::::|||
    383 rValThrPhe......GluGlySerGlyThrLeuThrLeuAsnAsnAsnI 398

   1232 TCAATCAAGGTGCTGGAGGATTATATTTCCAAGGAGATTTTACGGTC... 1278
         |||:::|||||||||||||||||||:::|||:::||||||:::     |||
    398 leAspGlnGlyAlaGlyGlyLeuPhePheGluGlyAspTyrGluValLys 414

   1279 TCGCCTGAAAATAACGAAACTTGGCAAGGCGCGGGCGTTCATATCAGTGA 1328
         :::     :::::::|||     ||||||:::||||||||||||     ::::::||
    415 GlyThrSerAspAsnThrThrTrpLysGlyAlaGlyValSerValAlaGl 431

   1329 AGACAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAAACGACCGCCTGT 1378
         |     :::||||||||||||||||||||::::::            |||||||||:
    431 uGlyLysThrValThrTrpLysValHisAsnProGlnTyrAspArgLeuA 448
```

```
1379 CCAAAATCGGCAAAGGCACGCTGCACGTTCAAGCCAAAGGGGAAAACCAA 1428
      ::||||||||||||||||||||        |||::::::    |||::|||:::
 448 laLysIleGlyLysGlyThrLeuIleValGluGlyThrGlyAspAsnLys 464

1429 GGCTCGATCAGCGTGGGCGACGGTACAGTCATTTGGATCAGCAGGCAGA 1478
      |||||:::::||||||||||||||||||||||||    ||||||::::
 465 GlySerLeuLysValGlyAspGlyThrValIleLeuLysGlnGlnThrAs 481

1479 CGATAAAGGCAAAAAACAAGCCTTTAGTGAAATCGGCTTGGTCAGCGGCA 1528
      :    :::|||     ::::::|||||||:::::::::|||:::||||||||||
 481 nGlySerGly...GlnHisAlaPheAlaSerValGlyIleValSerGlyA 497

1529 GGGGTACGGTGCAACTGAATGCCGATAATCAGTTCAACCCCGACAAACTC 1578
      ||:::|||:::    ||||||    |||:::|||     :::|||:::::::::
 497 rgSerThrLeuValLeuAsnAspAspLysGlnValAspProAsnSerIle 513

1579 TATTTCGGCTTTCGCGGCGGACGTTTGGATTTAAACGGGCATTCGCTTTC 1628
      ||||||||||||||||||||||||||||||||||||:::||||||:::
 514 TyrPheGlyPheArgGlyGlyArgLeuAspLeuAsnGlyAsnSerLeuTh 530

1629 GTTCCACCGTATTCAAAATACCGATGAAGGGGCGATGATTGTCAACCACA 1678
      :|||     :::|||:::|||     |||:::||||||    :::|||||||||
 530 rPheAspHisIleArgAsnIleAspAspGlyAlaArgLeuValAsnHisA 547

1679 ATCAAGACAAAGAATCCACCGTTACCATTACAGGC.............. 1713
      ||:::    :::    |||::::::||||||||||||
 547 snMetThrAsnAlaSerAsnIleThrIleThrGlyGluSerLeuIleThr 563

1713 ................................................ 1713

 564 AspProAsnThrIleThrProTyrAsnIleAspAlaProAspGluAspAs 580

1713 ................................................ 1713

 580 nProTyrAlaPheArgArgIleLysAspGlyGlyGlnLeuTyrLeuAsnL 597

1713 ................................................ 1713

 597 euGluAsnTyrThrTyrTyrAlaLeuArgLysGlyAlaSerThrArgSer 613

1713 ................................................ 1713

 614 GluLeuProLysAsnSerGlyGluSerAsnGluAsnTrpLeuTyrMetGl 630

1714 ....AATAAAGATATTGCTACAACCGGCAATAACAACAGCTTGGATAGCA 1759
          ::::::|||    |||          :::        |||    ::::::::::
 630 yLysThrSerAspGluAlaLysArgAsnValMetAsnHisIleAsnAsnG 647

1760 AAAAAGAAATTGCCTACAACGGTTGGTTTGGCGAGAAAGATACGACCAAA 1809
      :::::       ::::::|||||||:::||||||||||::::::      |||
 647 luArgMetAsnGlyPheAsnGlyTyrPheGlyGluGluGlu...GlyLys 662

1810 ACGAACGGGCGGCTCAACCTTGTTTACCAGCCCGCCGCAGAAGACCGCAC 1859
      :::||||||:::||||||:::::::::::::         :::|||::::::
 663 AsnAsnGlyAsnLeuAsnValThrPheLysGlyLysSerGluGlnAsnAr 679

1860 CCTGCTGCTTTCCGGCGGAACAAATTTAAACGGCAACATCACGCAAACAA 1909
      :::||||||:::||||||||||||||||||||||::::::|||          :
 679 gPheLeuLeuThrGlyGlyThrAsnLeuAsnGlyAspLeuThrValGluL 696

1910 ACGGCAAACTGTTTTTCAGCGGCAGACCAACACCGCACGCCTACAATCAT 1959
      ::|||      ||||||:::||||||||||||||||||||||||      :::
 696 ysGlyThrLeuPheLeuSerGlyArgProThrProHisAlaArgAspIle 712

1960 TTAAACGACCATTGGTCGCAAAAAGAGGGC......ATTCCTCGCGGGGA 2003
        :::          ::::::|||:::                ::::::||
 713 AlaGlyIleSerSerThrLysLysAspProHisPheAlaGluAsnAsnGl 729

2004 AATCGTGTGGGACAACGACTGGATCAACCGCACATTTAAAGCGGAAAACT 2053
      |:::|||     :::::|||||||||||||||:::|||||||||     ::::
 729 uValValValGluAspAspTrpIleAsnArgAsnPheLysAlaThrThrM 746

2054 TCCAAATTAAAGGCGGACAGGCGGTGGTTTCC...CGCAATGTTGCCAAA 2100
      ::::::::     |||:::     ::::::     |||     ||||||||||||:::
 746 etAsnValThrGlyAsnAlaSerLeuTyrSerGlyArgAsnValAlaAsn 762

2101 GTGAAAGGCGATTGGCATTTGAGCAATCACGCCCAAGCAGTTTTTGGTGT 2150
      :::     ::::::               ||||||     |||||||:::     :::|||
 763 IleThrSerAsnIleThrAlaSerAsnLysAlaGlnValHisIleGlyTy 779

2151 CGCACCGCATCAAAGCCACACAATCTGTACACGTTCGGACTGGACGGGTC 2200
                :::     |||:::|||:::|||||||||:::||||||
 779 r......LysThrGlyAspThrValCysValArgSerAspTyrThrGlyT 794

2201 TGACAAATTGTGTCGAAAAAACCATTACCGACGATAAAGTGATTGCTTCA 2250
        ::::::|||:::          ::::::     ||||||::::::     |||
 794 yrValThrCysThrThrAspLysLeuSer...AspLysAlaLeuAsnSer 809

2251 TTGACTAAGACCGACATCAGCGGCAATGTCGATCTTGCCGATCACGCTCA 2300
      ::::::     |||::::::     |||||||||:::|||::::::     |||::
 810 PheAsnProThrAsnLeuArgGlyAsnValAsnLeuThrGluSerAlaAs 826

2301 TTTAAATCTCACAGGGCTTGCCACACTCAACGGCAATCTTAGTGCAAATG 2350
      ::::     |||     |||     |||:::|||     |||::::::::::::::|
 826 nPheValLeu...GlyLysAlaAsnLeuPheGlyThrIleGlnSerArgG 842

2351 GCGATACACGTTATACAGTCAGCCACAACGCCACCCAAAACGGCAACCTT 2400
      ||:::::::::     :::::::::|||
 842 lyAsnSerGlnValArgLeuThrGluAsn.................... 851

2401 AGCCTCGTGGGCAATGCCCAAGCAACATTTAATCAAGCCACATTAAACGG 2450

 851 ................................................ 851

2451 CAACACATCGGCTTCGGGCAATGCTTCATTTAATCTAAGCGACCACGCCG 2500

 851 ................................................ 851

2501 TACAAAACGGCAGTCTGACGCTTTCCGGCAACGCTAAGGCAAACGTAAGC 2550

 851 ................................................ 851

2551 CATTCCGCACTCAACGGTAATGTCTCCCTAGCCGATAAGGCAGTATTCCA 2600

 851 ................................................ 851

2601 TTTTGAAAGCAGCCGCTTTACCGGACAAATCAGCGGCGGCAAGGATACGG 2650

 851 ................................................ 851

2651 CATTACACTTAAAAGACAGCGAATGGACGCTGCCGTCAGGCACGGAATTA 2700
                          |||:::|||     |||     ::::::::::::::
 852 .................SerHisTrpHisLeuThrGlyAsnSerAspVal 862

2701 GGCAATTTAAACCTTGACAACGCCACCATTACACTCAATTCCGCCTATCG 2750
        :::|||:::|||     |||:::     |||     ||||||||||||     ::
 863 HisGlnLeuAspLeuAlaAsnGlyHisIleHisLeuAsnSerAlaAspAs 879

2751 CCACGATGCGGCAGGGGCGCAAACCGGCAGTGCGACAGATGCGCCGCGCC 2800
      :     :::
 879 nSerAsn......................................... 881

2801 GCCGTTCGCGCCGTTCGCGCCGTTCCCTATTATCCGTTACACCGCCAACT 2850
                                    :::||||||
 882 ..............................AsnValThr......... 884

2851 TCGGTAGAATCCCGTTTCAACACGCTGACGGTAAACGGCAAATTGAACGG 2900
                ::::::||||||||||||||||||     :::|||:::||
 885 ...........LysTyrAsnThrLeuThrValAsn...SerLeuSerGl 896

2901 TCAGGGAACATTCCGCTTTATGTCGGAACTCTTCGGCTACCGCAGCGACA 2950
      |:::|||:::|||     ::::::::::::|||     :::     ::::::|||
 896 yAsnGlySerPheTyrTyrLeuThrAspLeuSerAsnLysGlnGlyAspL 913

2951 AATTGAAGCTGGCGGAAAGTTCCGAAGGCACTTACACCTTGGCGGTCAAC 3000
```

```
           ||:::   :::::::::|||:::   |||::::::||||||   |||
 913 ysValValValThrLysSerAlaThrGlyAsnPheThrLeuGlnValAla 929

3001 AATACCGGCAACGAACCTGCAAGCCTCGAACAATTGACGGTAGTGGAAGG 3050
     :::      :::||||||  :::   ::::::||||||:::   :::::
 930 AspLysThrGlyGluPro...AsnHisAsnGluLeuThrLeuPheAspAl 945

3051 AAAAGACAACAAACCGCTGTCCGAAAACCTTAATTTCACCCTGCAAAACG 3100
     :       :::|||      ::::::||||||  :::|||   ::::
 945 a......SerLysAlaGlnArgAspHisLeuAsnValSerLeuValGlyA 960

3101 AACACGTCGATGCCGGCGCGTGGCGTTACCAACTCATCCGCAAAGACGGC 3150
     ::    ||||||  ||||||||:::|||:::|||   :::   :::|||
 960 snThrValAspLeuGlyAlaTrpLysTyrLysLeuArgAsnValAsnGly 976

3151 GAGTTCCGCCTGCATAATCCGGTCAAAGAACAA................. 3183
     ::::::   |||:::||||||      |||:::
 977 ArgTyrAspLeuTyrAsnProGluValGluLysArgAsnGlnThrValAs 993

3183 .................................................. 3183

 993 pThrThrAsnIleThrThrProAsnAsnIleGlnAlaAspValProSerV 1010

3184 .....GAGCTTTCCGACAAACTCGGCAAG.................... 3207
          :::   ::::::::::::::::::
1010 alProSerAsnAsnGluGluIleAlaArgValAspGluAlaProValPro 1026

3208 .........................................GCAGAAGC 3215
                                              ||||||
1027 ProProAlaProAlaThrProSerGluThrThrGluThrValAlaGluAs 1043

3216 CAAAAAACAGGCGGAAAAA.........GACAACGCGCAAAGCCTTGACG 3256
       :::||||||   :::|||          |||   |||:::
1043 nSerLysGlnGluSerLysThrValGluLysAsnGluGlnAspAlaThrG 1060

3257 CGCTGATTGCGGCCGGGCGCGATGCCGTCGAA.................. 3288
                |||    :::|||::::::::::::
1060 luThrThrAlaGlnAsnArgGluValAlaLysGluAlaLysSerAsnVal 1076

3289 AAGACAGAAAGCGTTGCCGAACCGGCCCGGCAGGCAGGCGGGGAAAATGT 3338
     |||:::::::::   ::::::   :::   |||:::|||:::|||:::
1077 LysAlaAsnThrGlnThrAsnGluValAlaGlnSerGlySerGluThrLy 1093

3339 CGGCATTATGCAGGCGGAG..........................GAAG 3361
                                                 ||||
             :::   :::|||
1093 sGluThrGlnThrThrGluThrLysGluThrAlaThrValGluLysGluG 1110

3362 AGAAAAAACGGGTGCAGGCGGATAAAGACACCGCCTTGGCGAAA...... 3405
     |||||   :::|||:::::::::|||          :::   |||
1110 luLysAlaLysValGluThrGluLysThrGlnGluValProLysValThr 1126

3406 ........................................CAGCGCGAAGC 3416
                                              |||   ::|||
1127 SerGlnValSerProLysGlnGluGlnSerGluThrValGlnProGlnAl 1143

3417 GGAA.............................................. 3420
     ||||
1143 aGluProAlaArgGluAsnAspProThrValAsnIleLysGluProGlnS 1160

3420 .................................................. 3420

1160 erGlnThrAsnThrThrAlaAspThrGluGlnProAlaLysGluThrSer 1176

3420 .................................................. 3420

1177 SerAsnValGluGlnProValThrGluSerThrThrValAsnThrGlyAs 1193

3421 ........................ACCCGGCCGGCTACCACCGCCTTCC 3445
                             |||   ||||||||||||     |
1193 nSerValValGluAsnProGluAsnThrThrProAlaThrThrGln...P 1209

3446 CC.................................CGCGCCCGCCGC 3459
     ||                                 |||   ||||||

1209 roThrValAsnSerGluSerSerAsnLysProLysAsnArgHisArgArg 1225

3460 GCCCGCCGGGATTTGCCG...CAACTGCAACCCCAACCGCAGCCCCAACC 3506
     :::   |||::::::|||   :::::::::|||          :::
1226 SerValArgSerValProHisAsnValGluProAlaThrThrSerSerAs 1242

3507 GCAGCGC.................GACCTGATCAGCCGTTATGCCAATA 3538
      :::|||                 ||||||   |||      :::|||:
1242 nAspArgSerThrValAlaLeuCysAspLeuThrSerThrAsnThrAsnA 1259

3539 GCGGTTTGAGTGAATTTTCCGCC..............ACGCTCAACAGC 3573
     ::    ||||||:::        |||           :::||||||
1259 laValLeuSerAspAlaArgAlaLysAlaGlnPheValAlaLeuAsnVal 1275

3574 GTTTTCGCCGTACAGGACGAATTAGACCGCGTA...TTTGCCGAAGACCG 3620
          ||||||::::::::::::::::::::::   :::   ::::::
1276 GlyLysAlaValSerGlnHisIleSerGlnLeuGluMetAsnAsnGluGl 1292

3621 CCGCAACGCCGTTTGGACAAGCGGCATCCGGGACACCAAACACTACCGTT 3670
      :::        ||||||:::|||:::           :::|||:::|||    |
1292 yGlnTyrAsnValTrpValSerAsnThrSerMetAsnLysAsnTyrSerS 1309

3671 CGCAAGATTTCCGCGCCTACCGCCAACAAACCGACCTGCGCCAAATCGGT 3720
     ||::::::::|||   :::   :::::::::            |||:::|||
1309 erSerGlnTyrArgArgPheSerSerLysSerThrGlnThrGlnLeuGly 1325

3721 ATGCAGAAAAACCTCGGCAGCGGG...CGCGTCGGCATCCTGTTTTCGCA 3767
       ::::::::::::::::::::      ::::::|||   :::|||:::::
1326 TrpAspGlnThrIleSerAsnAsnValGlnLeuGlyGlyValPheThrTy 1342

3768 CAACCGGACCGAAAACACCTTCGACGACGGCATCGGCAACTCGGCACGGC 3817
     :     |||::::::|||:::||||||   :::   ::::::::::::     |
1342 rValArgAsnSerAsnAsnPheAspLysAlaThrSerLysAsnThr...L 1358

3818 TTGCCCACGGCGCCGTTTTCGGGCAATACGGCATCGAC...AGGTTCTAC 3864
     ||||||:::            :::::::::|||      |||   ::::::|||
1358 euAlaGlnValAsnPheTyrSerLysTyrTyrAlaAspAsnHisTrpTyr 1374

3865 ATCGGCATCAGCGCGGGCGCGGGTTTAGCAGCGGCAGCCTTTCAGACGG 3914
     :::|||||||:::   |||   |||    :::::::::|||:::   ::
1375 LeuGlyIleAspLeuGlyTyrGlyLysPheGlnSerLysLeuGlnThrAs 1391

3915 CATCGGAGGCAAAATCCGCCGCCGCGTGCTGCATTACGGCATTCAGGCAC 3964
     :    ::::::|||:::   |||::::::   ::::::|||:::   |||
1391 nHisAsnAlaLysPheAlaArgHisThrAlaGlnPheGlyLeuThrAlaG 1408

3965 GATACCGCGCCGGTTTCGGCGGATTCGGCATCGAACCGCACATCGGCGCA 4014
               ::::::|||:::|||||||||   |||   :::|||:::
1408 lyLysAlaPheAsnLeuGlyAsnPheGlyIleThrProIleValGlyVal 1424

4015 ACGCGCTATTTCGTCCAAAAAGCGGATTACCGCTACGAAAACGTCAATAT 4064
               :::::::::::|||||||:::         ::::::::::::||
1425 ArgTyrSerTyrLeuSerAsnAlaAspPheAlaLeuAspGlnAlaArgIl 1441

4065 CGCCACCCCCGGCCTTGCATTCAACCGCTACCGCGCGGGCATTAAGGCAG 4114
     |    :::      ::::::   :::            |||   :::      :
1441 eLysValAsnProIleSerValLysThrAlaPheAlaGlnValAspLeuS 1458

4115 ATTATTCATTCAAACCGGCGCAACACATTTCCATCACGCCTTATTTGAGC 4164
     :::|||::::::         ::::::   :::|||:::||||||   ||||||
1458 erTyrThrTyrHisLeuGlyGlu...PheSerValThrProIleLeuSer 1473

4165 CTGTCCTATACCGATGCCGCTTCGGGCAAAGTCCGAACACGCGTCAATAC 4214
          |||   ||||||   :::|||:::   :::   :::||||||
1474 AlaArgTyr...AspAlaAsnGlnGlySerGlyLysIleAsnValAsnGl 1489

4215 CGCCGTATTGGCTCAGGATTTCGGCAAAACCCGCAGTGCGGAATGGGGCG 4264
            :::|||   :::      :::   ::::::   :::   |||:
1489 yTyrAspPheAlaTyrAsnValGluAsnGlnGlnGlnTyrAsnAlaGlyL 1506

4265 TAAACGCCGAAATCAAAGGTTTCACGCTGTCCCTCCACGCTGCCGCCGCC 4314
     :::::   :::         :::      |||||||||   ::::::   :::
1506 euLysLeuLysTyrHisAsnValLysLeuSerLeuIleGlyGlyLeuThr 1522
```

```
4315 AAAGGCCCGCAACTGGAAGCGCAACACAGCGCGGGCATCAAATTAGGCTA 4364
      |||:::   |||   |||   |||   :::|||   :::||||||::::
1523 LysAlaLysGlnAlaGluLysGlnLysThrAlaGluLeuLysLeuSerPh 1539

4365 C 4365
     :
1539 e 1539

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY75564

seq_documentation_block:
ID   AAY75564 standard; Protein; 1431 AA.
XX
AC   AAY75564;
XX
DT   21-MAR-2000  (first entry)
XX
DE   Neisseria meningitidis ORF 759 protein sequence SEQ ID NO:2602.
XX
KW   Neisseria meningitidis; Neisseria gonorrheae; antigen; vaccine;
KW   antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW   antibacterial; gene therapy.
XX
OS   Neisseria meningitidis.
XX
PN   WO9957280-A2.
XX
PD   11-NOV-1999.
XX
PF   30-APR-1999;   99WO-US09346.
XX
PR   01-MAY-1998;   98US-0083758.
PR   31-JUL-1998;   98US-0094869.
PR   02-SEP-1998;   98US-0098994.
PR   02-SEP-1998;   98US-0099062.
PR   09-OCT-1998;   98US-0103749.
PR   09-OCT-1998;   98US-0103794.
PR   09-OCT-1998;   98US-0103796.
PR   25-FEB-1999;   99US-0121528.
XX
PA   (CHIR ) CHIRON CORP.
PA   (GENO-) INST GENOMIC RES.
XX
PI   Fraser C,  Galeotti C,  Grandi G,  Hickey E,  Masignani V,  Mora M;
PI   Petersen J,  Pizza M,  Rappuoli R,  Ratti G,  Scalato E,  Scarselli M;
PI   Tettelin H,  Venter JC;
XX
DR   WPI; 2000-062150/05.
DR   N-PSDB; AAZ54326.
XX
PT   Novel Neisserial polypeptides predicted to be useful antigens for
PT   vaccines and diagnostics  -
XX
PS   Claim 2; Page 1234; 1453pp; English.
XX
CC   AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
CC   represent novel Neisseria meningitis and N. gonorrheae polynucleotides
CC   and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ55473 represent
CC   PCR primers used in the exemplification of the present invention. The
CC   polypeptides, the polynucleotides, antibodies and compositions of
CC   the invention can be used as vaccines, as diagnostic reagents, and as
CC   immunogenic compositions. The polypeptides can be used in the
CC   manufacture of medicaments for treating or preventing infection due to
CC   Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
CC   presence of Neisseria bacteria, or to raise antibodies. They may also
CC   be used to screen for agonists or antagonists, which may themselves
CC   have use as antibacterial agents. The polynucleotides of the invention
CC   may also be used in gene therapy protocols.
XX
SQ   Sequence   1431 AA;

alignment_scores:
          Quality: 1547.00            Length:   1669
            Ratio:  1.809               Gaps:     50
 Percent Similarity: 51.228   Percent Identity: 27.741

alignment_block:
 US-09-303-518D-649 x AAY75564  ..

 Align seg 1/1  to: AAY75564  from: 1  to: 1431

  55 GGCCGCATCCGCTTCTCGCCTGCTTACTTAGCCATATGCCTGTCGTTCGG 104
      |||   :::      |||||||||||   :::|||      :::
  17 GlyLeuPheAlaValSerProAlaTyrSerSerIleValArgAsn..... 31

 105 CATTCTTCCCCAAGCCTGGGCGGGACACACTTATTTCGGCATCAACTACC 154
                                           :::::::||||
  32 ..................................AspValAspTyrG 36

 155 AATACTATCGCGACTTTGCCGAAAATAAAGGCAAGTTTGCAGTCGGGGCG 204
     |||||:::||||||||||||||||||||||||   |||:::|||||||||
  36 lnTyrPheArgAspPheAlaGluAsnLysGlyAlaPheThrValGlyAla 52

 205 AAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTGGTCGGCAAATCAAT 254
     ::::::|||::::::   :::|||:::|||::::::::::|||:::   ::
  53 SerAsnIleSerIleGlnAspLysGlnGlyLysIleLeuGlyArgValLe 69

 255 GACAAAAGCCCCGATGATTGATTTTTCTGTGGTGTCGCGT...AACGGCG 301
     ::::       ||||||  ||||||  |||   :::|||   :::::::
  69 uAsnGlyIleProMetProAspPheArgValSerAsnArgGlnThrAlaI 86

 302 TGGCGGCATTGGTGGGCGATCAATATATTGTGAGCGTGGCACATAACGGC 351
     ::|||:::||||||          |||||||:::   ||||||   ||||||
  86 leAlaThrLeuValHisProGlnTyrValAsnSerValLysHisAsnVal 102

 352 GGCTATAACAACGTTGATTTTGGTGCGGAAGGAAGAAATCCCGATCAACA 401
     ||||||:::::::::::||||||   :::   :::||||||::::::::
 103 GlyTyrGlySerIleGlnPheGlyAsnAspThrGlnAsnProGluGlu Gl 119

 402 TCGTTTTACTTATAAAATTGTGAAACGGAATAATTATAAAGCAGGGACTA 451
     :   :::|||||||::::::|||:::||||||
 119 nAlaTyrThrTyrArgLeuValSerArgAsn.................. 129

 452 AAGGCCATCCTTATGGCGGCGATTATCATATGCCGCGTTTGCATAAATTT 501
         ||||||         |||||||||:::||||||||||:::|||:::
 130 ..ProHisPro...AspTyrAspTyrHisLeuProArgLeuAsnLysLeu 144

 502 GTCACAGATGCAGAACCTGTTGAAATGACCAGT................. 534
     |||||||:::   :::|||:::   ::::::|||
 145 ValThrGluIleSerProThrAlaLeuSerSerValProLeuLeuGlyAs 161

 535 ....TATATGGATGGGCGGAAATATATCGATCAAAATAATTACCCTGACC 580
                  ::::::   |||:::|||   :::::::::|||
 161 nGlyGlnProLysAlaAsnAlaTyrLeuAspThrAspArgPheProTyrP 178

 581 GTGTTCGTATTGGGGCAGGCAGGCAATATTGGCGATCTGATGAAGATGAG 630
      ||||||:::|||:::|||   |||      |||:::   :::
 178 heValArgLeuGlySerGlyThrGlnGlnValArgLysAlaAspGly... 193

 631 CCCAATAACCGCGAAAGTTCATATCATATTGCAAGTGCGTATTCTTGGCT 680
                      :::   :::   |||   ||||||::::::||
 194 ..............ThrArgThrArgThrAlaProAlaTyrGlnTyrLe 205

 681 CGTTGGTGGCAATACCTTTGCACAAAATGGATCAGGTGGTGGCACAGTCA 730
     |:::|||||||:::   :::         |||      :::      :::
 205 uThrGlyGlyThrProLeuLysValLeuGlyPheGlnAsnHisGlyLeuL 222

 731 ACTTAGCTAGTGAAAAAATTAAACATAGCCCATATGGTTTTTACCAACA 780
       :::|||:::   ::::::      :::|||         |||   |||
 222 euValGlyGly...SerLeuThrAspGlnPro.........LeuAsnThr 234

 781 GGAGGCTCATTTGGCGACAGTGGCTCACCAATGTTTATCTATGATGCCCA 830
       :::         ||||||||||||||||||:::|||   :::|||   ::
 235 TyrAlaIleAlaGlyAspSerGlySerProLeuPheAlaPheAspLysHi 251
```

```
 831 AAAGCAAAAGTGGTTAATTAATGGGGTATTGCAAACGGGCAACCCCTATA 880
     :::::::::|||::::::    ||||||||||:::|||         |||
 251 sGluAsnArgTrpValLeuAlaGlyValLeuSerThr.........TyrA 265

 881 TAGGAAAAAGCAATGGCTTCCAG......CTGGTTCGTAAAGATTGGTTC 924
      |||    :::|||   |||:::       :::|||   :::      |||
 265 laGlyPheAspAsnPhePheAsnLysTyrIleValThrGlnProGluPhe 281

 925 TATGATGAAATCTTTGCTGGAGATACCCATTCAGTATTCTACGAACCACG 974
        :::   :::                             ||||||   ||
 282 IleArgSerThrIleArgGln..................TyrGluThrAr 292

 975 T..............CAAAATGGGAAATACTCTTTTAACGACGATAATA 1009
     |                 :::::::::           ::::::|||::::::|
 292 gLeuAspValGlyLeuThrThrAsnGluLeuIleTrpArgAspAsnGlyA 309

1010 ATGGCACAGGAAAAATCAATGCCAAACATGAACACAATTCTCTGCCTAAT 1059
     |||||:::::   :::::::::   :::|||:::   :::||||||
 309 snGlyAsnSerThrLeuGlnGlyLeuAsnGluArgIleThrLeuPro... 324

1060 AGATTAAAAACACGAACCGTTCAATTGTTTAATGTTCTTTATCCGAGAC 1109
                               :::   |||   |||||||:::
 325 ........................IleAlaAsnProSerLeuAla..... 331

1110 AGCAAGAGAACCTGTTTATCATGCTGCAGGTGGTGTCAACAGTTATCGAC 1159
                                                    |
 332 ..............................................P 332

1160 CCAGACTGAATAATGGAGAAAATATTTCCTTTATTGACGAAGGAAAA... 1206
      ||:::   |||:::::::::::::::        |||    ||||||
 332 roGln...AsnAspSerArgHisMetProSerGluAspAlaGlyLysThr 347

1207 ..........................GGCGAATTGATACTTACCAGCAA 1229
                               |||:::|||::::::||
 348 LeuIleLeuSerSerArgPheAspAsnLysThrLeuMetLeuAlaAspAs 364

1230 CATCAATCAAGGTGCTGGAGGATTATATTTCCAAGGAGATTTTACGGTCT 1279
      ||||||||||||||||||:::|||   |||:::::::::|||||||||
 364 nIleAsnGlnGlyAlaGlyAlaLeuGlnPheAspSerAsnPheThrValV 381

1280 CGCCTGAAAATAACGAAACTTGGCAAGGCGCGGGCGTTCATATCAGTGAA 1329
        :::|||:::    |||||||||||||||||||||    :::::::::
 381 alGlyLysAsnHis...ThrTrpGlnGlyAlaGlyValIleValAlaAsp 396

1330 GACAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAAACGACCGCCTGTC 1379
        :::   |||   |||:::|||::::::       :::||||||||||
 397 GlyLysArgValPheTrpGlnValSerAsnProLysGlyAspArgLeuSe 413

1380 CAAAATCGGCAAAGGCACGCTGCACGTTCAAGCCAAAGGGGAAAACCAAG 1429
      ||||:::|||   |||||||||   ::::::::::::|||   |||||||
 413 rLysLeuGlyAlaGlyThrLeuIleAlaAsnGlyGlnGlyIleAsnGlnG 430

1430 GCTCGATCAGCGTGGGCGACGGTACAGTCATTTTGGATCAGCAGGCAGAC 1479
      ||:::||||||:::|||:::|||||||||||:::|||    |||:::|||
 430 lyAspIleSerIleGlyGluGlyThrValValLeuAlaGlnLysAlaAla 446

1480 GATAAAGGCAAAAAACAAGCCTTTAGTGAAATCGGCTTGGTCAGCGGCAG 1529
      :::    |||:::|||||||||||||||:::::::::|||::::::||||||
 447 SerAspGlySerLysGlnAlaPheAsnGlnValGlyIleThrSerGlyAr 463

1530 GGGTACGGTGCAACTGAATGCCGATAATCAGTTCAACCCCGACAAACTCT 1579
      ||||||||:::   |||         ::::::|||:::::::|||:::::::||||
 463 gGlyThrAlaValLeuAlaAspSerGlnGlnIleLysProGluAsnLeuT 480

1580 ATTTCGGCTTTCGCGGCGGACGTTTGGATTTAAACGGGCATTCGCTTTCG 1629
      ||||||||||||||||||||||||||||||||||||||:::::::|||:::
 480 yrPheGlyPheArgGlyGlyArgLeuAspLeuAsnGlyAsnAsnLeuAla 496

1630 TTCCACCGTATTCAAAATACCGATGAAGGGGCGATGATTGTCAACCACAA 1679
      |||   :::|||:::::::::|||   ||||||:::|||||||||||||||
 497 PheThrHisIleArgHisAlaAspGlyGlyAlaGlnIleValAsnHisAs 513

1680 TCAAGACAAAGAATCCACCGTTACCATTACAGGCAATAAAGATATTGCTA 1729
     |   |||:::   :::|||:::|||:::|||||||||         ::::::
 513 nProAspGlnAlaAlaThrLeuThrLeuThrGlyAsnProValLeuSerP 530

1730 CAACC............................................. 1734

 530 roGluHisValGluTrpValGlnTrpGlyAsnArgProGlnGlyAsnAla 546

1734 .................................................. 1734

 547 AlaValTyrGluTyrIleAsnProHisArgAsnArgArgThrAspTyrPh 563

1735 ...............GGCAAT............................ 1740
                    ||||||
 563 eIleLeuLysProGlyGlyAsnProArgGluPhePheProLeuAsnMetL 580

1741 ...........................AACAACAGCTTGGATAGCAAA 1761
                                ||||||     ::::::
 580 ysAsnSerThrSerTrpGlnPheIleGlyAsnAsnArgGlnGlnAlaAla 596

1762 AAAGAAATTGCC..............................TACAA 1778
     :::::::::|||                              ::::
 597 GluGlnValAlaGlnAlaGluAsnAlaArgProAspLeuIleThrPheGl 613

1779 CGGTTGGTTTGGCGAGAAAGATACGACC...................... 1806
      :|||:::::|||||||:::      |||
 613 yGlyTyrLeuGlyGluAsnAlaGlnThrGlyLysAlaAlaProSerTyrS 630

1806 .................................................. 1806

 630 erLysThrAsnGluAlaAlaIleGluLysThrArgHisIleAlaAsnAla 646

1807 .......................AAAACGAACGGGCGGCTCAACCTTGT 1832
                            :::   ||||||   |||||||||
 647 AlaValTyrGlyArgProGluTyrArgTyrAsnGlyAlaLeuAsnLeuHi 663

1833 TTACCAGCCCGCCGCAGAAGACCGCACCCTGCTGCTTTCCGGCGGAACAA 1882
      |||:::|||      |||   |||||||||||:::||||||     |
 663 sTyrArgProLysArgThrAspSerThrLeuLeuLeuAsnGlyGlyMetA 680

1883 ATTTAAACGGCAACATCACGCAAACAAACGGCAAACTGTTTTTCAGCGGC 1932
      ||||||||||||::::::           :::|||:::::::::   ||||||
 680 snLeuAsnGlyGluValLeuIleGluGlyGlyAsnMetIleValSerGly 696

1933 AGACCAACACCGCACGCCTACAATCATTTAAACGACCATTGGTCGCAAAA 1982
      ||||||:::|||||||||||||           ||||||   ::::::::
 697 ArgProValProHisAlaTyr............AspHisGlnAlaLysAr 709

1983 AGAGGGCATTCCTCGCGGGGAAATCGTGTGGGACAACGACTGGATCAACC 2032
      :|||                       |||   :::|||:::|||   :::
 709 gGluPro..................ValLeuGluAsnGluTrpThrAspG 720

2033 GCACATTTAAAGCGGAAAACTTCCAAATTAAAGGC...GGACAGGCGGTG 2079
        :::|||||||||   :::|||   :::::::::   ::::::   :::
 720 lySerPheLysAlaAlaArgPheThrLeuArgAsnHisAlaArgLeuThr 736

2080 GTTTCCCGCAATGTTGCCAAAGTGAAAGGCGAT.........TGGCATTT 2120
      :::::::|||||||::::|||   :::   ||||||         :::   ||
 737 AlaGlyArgAsnThrAlaHisLeuAspGlyAspIleThrAlaTyrAspLe 753

2121 GAGCAAT...CACGCCCAAGCAGTTTTTGGTGTCGCACCGCATCAAAGCC 2167
      |||||:::                  :::   |||   :::|||:::
 753 uSerGlyIleAspLeuGlyPheThrGlnGlyLysThrProGlu....... 767

2168 ACACAATCTGTACACGTTCGGACTGGACGGGTCTGACAAATTGTGTCGAA 2217
              |||   ||||||        :::|||   |||:::|||:::
 768 ........CysTyrArgSerTyrHisSerGlySerThrHisCysThrPro 781

2218 AAAACCATTACCGACGATAAAGTGATTGCTTCATTGACTAAGACCGACAT 2267
      :::::::::            :::          :::|||   |||::::::
 782 AsnAlaValLeuLysAlaGluAsnTyrArgAlaLeuProAlaThrGlnVa 798

2268 CAGCGGCAATGTCGATCTTGCCGATCACGCTCATTTAAATCTCACAGGGC 2317
```

```
        :  |||:::::  |||  |||::::::::|||:::|||  |||
 798 lArgGlyAspIleThrLeuAsnAspArgSerGluLeuArgLeu...GlyL 814

2318 TTGCCACACTCAACGGCAATCTTAGTGCAAATGGCGATACACGTTATACA 2367
       |||  |||  |||::::  |||
 814 ysAlaHisLeuTyrGlySerIleArgAla..................... 823

2368 GTCAGCCACAACGCCACCCAAAACGGCAACCTTAGCCTCGTGGGCAATGC 2417

 823 .................................................. 823

2418 CCAAGCAACATTTAATCAAGCCACATTAAACGGCAACACATCGGCTTCGG 2467

 823 .................................................. 823

2468 GCAATGCTTCATTTAATCTAAGCGACCACGCCGTACAAAACGGCAGTCTG 2517

 823 .................................................. 823

2518 ACGCTTTCCGGCAACGCTAAGGCAAACGTAAGCCATTCCGCACTCAACGG 2567

 823 .................................................. 823

2568 TAATGTCTCCCTAGCCGATAAGGCAGTATTCCATTTTGAAAGCAGCCGCT 2617

 823 .................................................. 823

2618 TTACCGGACAAATCAGCGGCGGCAAGGATACGGCATTACACTTAAAA... 2664
                         |||||||||||||||::::::::::
 824 ....................GlyLysAspThrAlaValArgMetGluAla 833

2665 GACAGCGAATGGACGCTGCCGTCAGGCACGGAATTAGGCAATTTAAACCT 2714
     ||||||:::|||||||||  ::::::::::::  |||  |||:::||
 834 AspSerAsnTrpThrLeuSerGlnSerSerHisThrGlyAlaLeuThrLe 850

2715 TGACAACGCCACCATTACACTCAATTCCGCCTATCGCCACGATGCGGCAG 2764
     ||||:::|||  |||||||||||||  :::  ::::::
 850 uAspGlyAlaGlnIleThrLeuAsnProAspPheAlaAsnAsn....... 864

2765 GGGCGCAAACCGGCAGTGCGACAGATGCGCCGCGCCGCCGTTCGCGCCGT 2814

 864 .................................................. 864

2815 TCGCGCCGTTCCCTATTATCCGTTACACCGCCAACTTCGGTAGAATCCCG 2864
                                   :::  ::::::||
 865 ..................................ThrHisAsnAsnAr 869

2865 TTTCAACACGCTGACGGTAAACGGCAAATTGAACGGTCAGGGAACATTCC 2914
     |||||||||||||||||||||||||||  |||:::|||  ||||||||||
 869 gPheAsnThrLeuThrValAsnGlyThrLeuAspGlyPheGlyThrPheA 886

2915 GCTTTATGTCGGAACTCTTC...GGCTACCGCAGCGACAAATTGAAGCTG 2961
     |||||::::::  :::       ::::::      |||||||||
 886 rgPheLeuThrGlyIleValArgLysGlnAsnAlaProProLeuLysLeu 902

2962 GCGGAAAGTTCCGAAGGCACTTACACCTTGGCGGTCAACAATACCGGCAA 3011
          :::|||:::|||::::::  :::  |||:::|||||||||::
 903 GluGlyAspSerArgGlyAlaPheGlnIleHisValLysAsnThrGlyGl 919

3012 CGAACCTGCAAGCCTCGAACAATTGACGGTAGTGGAAGGAAAAGACAACA 3061
     :||||||  :::  |||:::|||::::::||| :::  |||
 919 nGluProGlnThrThrGluSerLeuAlaLeuVal......SerLeuAsnP 934

3062 AACCGCTGTCCGAAAACCTTAATTTCACCCTGCAAAACGAACACGTCGAT 3111
             |||:::::  :::|||||||||||||||  ::::::|||
 934 roLysHisSerHisGlnAlaArgPheThrLeuGlnAsnGlyTyrAlaAsp 950

3112 GCCGGCGCGTGGCGTTACCAACTCATCCGCAAAGACGGCGAGTTCCGCCT 3161
       ||||||:::||||||  |||  ::::::::::::  :::  ||
 951 LeuGlyAlaTyrArgTyrIleLeuArgLysAsnAsnAsnGlyTyrSerLe 967

3162 GCATAATCCGGTCAAAGAACAAGAGCTTTCCGACAAACTCGGCAAGGCAG 3211
     |:::||||||:::||||||  ||||||:::  :::     :::||||
 967 uTyrAsnProLeuLysGluAlaGluLeuGlnIleGluAlaThrArgAlaG 984

3212 AA.........................................GCC 3216
     ||
 984 luHisGluArgAsnGlnGlnAlaTyrAsnGlnLeuGlnAlaThrAspIle 1000

3217 AAAAAACAGGCGGAAAAAGACAAC.......................... 3240
     ::::::|||::::::  |||:::
1001 SerArgGlnValGlnHisAspSerAspAlaThrArgGlnAlaLeuGlnAl 1017

3240 .................................................. 3240

1017 aTrpGlnAsnSerGlnThrGluLeuAlaArgIleAspSerGlnValGlnT 1034

3240 .................................................. 3240

1034 yrLeuSerAlaGlnLeuLysGlnThrAspProLeuThrGlyIleLeuThr 1050

3241 ...GCGCAAAGCCTTGACGCGCTG.......................... 3261
        ||||||:::|||  |||
1051 ArgAlaGlnAsnLeuCysAlaAlaGlnGlyTyrSerAlaAspIleCysAr 1067

3262 ....ATTGCGGCCGGGCGCGATGCC......................... 3282
        :::|||  :::  |||:::
1067 gGlnValAlaLysAlaAlaAspThrAsnAspLeuThrLeuPheGluThrG 1084

3283 ......................GTCGAAAAGACAGAAAGCGTTGCCGAA 3309
                           ||||||  :::||||||    :::
1084 luLeuAspThrTyrIleGluArgValGluMetAlaGluSerGluLeuAsp 1100

3310 CCGGCCCGGCAGGCAGGCGGGGAAAATGTCGGCATTATGCAGGCGGAGGA 3359
      ||||||||:::|||                        ||||||  ||
1101 LysAlaArgGlnGlyGlyAsp..............AlaGlnAlaValGl 1112

3360 AGAGAAAAAACGGGTGCAGGCGGATAAAGACACCGCCTTGGCGAAACAG. 3408
     |   ::::::::  :::  :::  |||::::::|||
1112 uThrAlaArgHisAlaTyrLeuAsnAlaLeuAsnArgLeuSerArgGlnI 1129

3409 ..CGCGAAGCGGAAACCCGGCCGGCTACCACCGCCTTCCCCCGCGCCCGC 3456
       ::::::  :::|||  |||  :::|||:::
1129 leHisSerLeuLysThrGlyValAlaGlyIleArgMetProAsnLeu... 1144

3457 CGCGCCCGCCGGGATTTGCCGCAACTGCAACCCCAACCGCAGCCCCAACC 3506

1144 .................................................. 1144

3507 GCAGCGCGACCTGATCAGCCGTTATGCCAATAGCGGTTTGAGTGAATTTT 3556
        :::|||||||||||||  ||||||:::::::::||||||  :
1145 ....AlaGluLeuIleSerArgSerAlaAsnThrAlaValSerGluGlnA 1160

3557 CCGCCACGCTCAACAGCGTTTTCGCCGTACAGGACGAATTAGACCGCGTA 3606
     ::|||  ::::::  :::  ::::::||||||
1160 laAlaTyrAsnThrGlyArgGlnGlnAlaGlyArgArgIleAspArgHis 1176

3607 TTTGCCGAAGACCGCCGCAACGCCGTTTGGACAAGCGGCATCCGGGACAC 3656
     :::::::::  :::::::::  :::|||  :::      :::
1177 LeuThrAspProGlnGlnGlnAsnIleTrpLeuGluThrGlyThrGlnGl 1193

3657 CAAACACTACCGTTCGCAAGATTTCCGCGCCTACCGCCAACAAACCGACC 3706
        |||:::|||  |||  |||:::|||  |||:::
1193 nThrAspTyrHisSerGlyThrHisArgProTyrGlnGlnThrThrAsnT 1210

3707 TGCGCCAAATCGGTATGCAGAAAAACCTCGGCAGCGGG...CGCGTCGGC 3753
        :::|||||||:::|||  ::::::  :::       ||||||
1210 yrAlaHisIleGlyIleGlnThrGlyIleThrAspArgLeuSerValGly 1226

3754 ATCCTGTTTTCGCACAACCGGACCGAAAACACCTTCGACGACGGCATCGG 3803
        :::::::::  :::||||||:::|||  ||||||:::|||::::
1227 ThrIleLeuThrAspGluArgThrAsnAsnArgPheAspGluGlyValSe 1243

3804 C......AACTCGGCACGGCTTGCCCACGGCGCCGTTTTCGGGCAATACG 3847
     :   |||  ::::::  ||||||   |||  |||:::  |
1243 rAlaArgAsnArgSerAsnGlyAlaHisLeuPheValLysGlyGluAsnG 1260
```

```
3848 GCATCGACAGGTTCTACATCGGCATCAGCGCGGGCGCGGGTTTTAGCAGC 3897
     ||             :::      :::|||      |||:::|||:::
1260 lyAla............LeuPheAlaAlaAlaAspLeuGlyTyrSerAsn 1272

3898 GGC......AGCCTTTCAGACGGCATCGGAGGCAAAATCCGCCGCCGCGT 3941
     :::         ::::::|||      |||:::   :::||||||::::
1273 SerArgThrArgPheThrAspTyrAspGlyAlaAlaValArgArgHisAl 1289

3942 GCTGCATTACGGCATTCAGGCACGATACCGCGCCGGTTTCGGCGGATTCG 3991
     :          ||||||::::::     :::              |||::::
1289 aTrpAspAlaGlyIleAsnThrGlyIleLysIleAsp...ThrGlyIleA 1305

3992 GCATCGAACCGCACATCGGC.....................GCAACGCGC 4020
     ::::::::|||:::    |||                    :::::::|||
1305 snLeuArgProTyrAlaGlyIleArgIleAsnArgSerAsnGlyAsnArg 1321

4021 TATTTCGTCCAAAAAGCGGATTACCGCTACGAAAACGTCAATATCGCCAC 4070
     |||   ::::::    |||:::                   |||    ::
1322 TyrValLeuAspGlyAlaGlu.....................IleAsnSe 1331

4071 CCCCGGCCTTGCATTCAACCGCTACCGCGCGGGCATTAAGGCAGATTATT 4120
     :|||:::            :::    ::::::|||||||||:::   |||    :
1331 rProAlaGlnIleGlnThrThrTrpHisAlaGlyIleArgLeuAspLysT 1348

4121 CATTCAAACCGGCGCAACACATTTCCATCACGCCTTATTTGAGCCTGTCC 4170
     ::   :::   :::|||       ::::::||||||   :::|||    :::
1348 hrValGluLeuGlyGln...AlaLysLeuThrProAlaPheSerSerAsp 1363

4171 TAT......ACCGATGCCGCTTCGGGCAAAGTCCGAACACGCGTCAATAC 4214
     |||        |||          ||||||::::::   :::    ||||||
1364 TyrTyrHisThrArgGlnAsnSerGlySerAlaLeuSer...ValAsnAs 1379

4215 CGCCGTATTGGCTCAGGATTTCGGCAAAACCCGCAGTGCGGAATGGGGCG 4264
          :::|||   |||:::    :::                        :
1379 pArgThrLeuLeuGlnGlnAlaAlaHisGlyThrLeuHisThrLeuGlnI 1396

4265 TAAACGCCGAAATCAAAGGTTTCACGCTGTCCCTCCACGCTGCCGCCGCC 4314
     ::::::|||       ||||||::::::    :::||||||||||||    :::
1396 leAspAlaGlyTyrLysGlyTrpAsnAlaLysLeuHisAlaAlaTyrGly 1412

4315 AAAGGCCCGCAACTGGAAGCGCAACACAGCGCGGGCATCAAATTAGGCTA 4364
     |||         :::             :::    :::||||||||||||:::|||||
1413 LysAspSerAsnThrAlaArgHisLysGlnAlaGlyIleLysIleGlyTy 1429

4365 CCGCTGG 4371
      |:::|||
1429 rAsnTrp 1431
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABB52745

```
seq_documentation_block:
ID   ABB52745 standard; Protein; 1371 AA.
XX
AC   ABB52745;
XX
DT   11-FEB-2002  (first entry)
XX
DE   Escherichia coli polypeptide SEQ ID NO 891.
XX
KW   Escherichia coli; B2/D+A-; antiinflammatory; antibacterial;
KW   immunosuppressive; extra-intestinal infection; phylogeny; meningitis;
KW   systemic infection; non-diarrhoeal infection; septicaemia;
KW   pyelonephritis; antibiotic resistance.
XX
OS   Escherichia coli.
XX
PN   WO200166572-A2.
XX
PD   13-SEP-2001.
XX
PF   12-MAR-2001; 2001WO-EP03445.
XX
PR   10-MAR-2000; 2000FR-0003145.
PR   02-FEB-2001; 2001FR-0001449.
XX
PA   (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
XX
PI   Bingen E,  Bonacorsi S,  Clermont O,  Nassif X,  Tinsley C;
XX
DR   WPI; 2001-550253/61.
XX
PT   A library of DNA fragments of Escherichia coli strains for the
PT   phylogenic determination of a given strain comprises polynucleotides of
PT   nature B2/D+ A- -
XX
PS   Example 6; Fig 6; 646pp; English.
XX
CC   The invention relates to a library of DNA fragments of Escherichia coli
CC   strains comprising polynucleotides (ABA88577-ABA88729 and ABA89533)
CC   and encoded proteins (ABB52459-ABB52919 and ABB52954-ABB53094) of nature
CC   B2/D+A-. The polynucleotides have potential antiinflammatory,
CC   antibacterial and immunosuppressive activity as part of pharmaceutical
CC   compositions used to treat, palliate or prevent extra-intestinal E. coli
CC   infections. The polypeptides are useful for determining the phylogenic
CC   group of a given E. coli strain. These polypeptides can detect and treat
CC   an undesired development of E. coli, particularly an extra-intestinal
CC   infection that include systemic and non-diarrhoeal infections such as
CC   septicaemia, pyelonephritis and meningitis this is particularly
CC   advantageous as bacterial resistance is increasing with the more
CC   frequent use of broad spectrum antibiotics.
XX
SQ   Sequence   1371 AA;

alignment_scores:
           Quality: 1071.00              Length:   1553
             Ratio:  1.370                 Gaps:     64
 Percent Similarity:  50.354    Percent Identity: 25.821

alignment_block:
 US-09-303-518D-649 x ABB52745   ..

 Align seg 1/1  to: ABB52745  from: 1  to: 1371

 112 CCCCAA...GCCTGGGCGGGACACACTTATTTCGGCATCAACTACCAATA 158
      ||||||    :::    ||||||    :::             |||    ||||||
  50 ProGlnIleSerGlnAlaGlyIleValArgSerAspIleAlaTyrGlnIl 66

 159 CTATCGCGACTTTGCCGAAAATAAAGGCAAGTTTGCAGTCGGGGCGAAAG 208
       |||||||||||||||||||||||||||    |||:::    ||||||    |
  66 eTyrArgAspPheAlaGluAsnLysGlyLeuPheValProGlyAlaThrA 83

 209 ATATTGAGGTTTACAACAAAAAAGGGGAGTTGGTCGGCAAATCAATGACA 258
     |||||    ||||||:::|||    |||:::|||||||||:::    :::
  83 spIleProValTyrAspLysAspGlyLysLeuValGlyArg...LeuAsp 98

 259 AAAGCCCCGATGATTGATTTTTCTGTGGTGTCGCGTAACGGCGTGGCGGC 308
     ||||||||||||    |||||||||    ||||||    ||||||||||||::
  99 LysAlaProMetAlaAspPheSerSerValSerSerAsnGlyValAlaTh 115

 309 ATTGGTGGGCGATCAATATATTGTGAGCGTGGCACATAACGGCGGCTATA 358
     :||||||:::    ||||||||||||||||||    |||||||||||||||:
 115 rLeuValSerProGlnTyrIleValSerValLysHisAsnGlyGlyTyrG 132

 359 ACAACGTTGATTTTGGTGCGGAAGGAAGAAATCCCGATCAACATCGTTTT 408
     ::::::|||:::||||||    :::|||:::|||
 132 lnSerValSerPheGly...AsnGlyLysAsn.................. 141

 409 ACTTATAAAATTGTGAAACGGAATAATTATAAAGCAGGGACTAAAGGCCA 458
     ||||||::::::|||    |||||||||:::::::::
 142 ThrTyrSerLeuValAspArgAsnAsnHisSerSer.............. 153

 459 TCCTTATGGCGGCGATTATCATATGCCGCGTTTGCATAAATTTGTCACAG 508
                    |||:::|||    |||||||||:::|||:::||||||:
 154 .........ValAspPheHisAlaProArgLeuAsnLysLeuValThrG 167
```

```
 509 ATGCAGAACCTGTTGAAATGACCAGTTATATGGATGGGCGGAAA...... 552
 167 luValIleProSerAlaIleThrSer......GluGlyThrLysAlaAsn 181

 553 ...TATATCGATCAAAATAATTACCCTGACCGTGTTCGTATTGGGGCAGG 599
 182 AlaTyrLysAspThrGluArgTyrThrAlaPheTyrArgValGlySerGl 198

 600 CAGGCAATATTGGCGATCTGATGAAGATGAGCCCAATAACCGCGAAAGTT 649
 198 yThrGlnTyrThrLys...AspLysAsp..............GlyAsnL 209

 650 CATATCATATTGCAAGTGCGTATTCTTGGCTCGTTGGTGGCAATACCTTT 699
 209 euValLysValAlaGlyGlyTyrAlaPheLysThrGlyGlyThrThr... 224

 700 GCACAAAATGGATCAGGTGGTGGCACAGTCAACTTAGGTAGTGAAAAAAT 749
 225 ....................GlyValProLeuIleSerAspAlaThrIl 234

 750 TAAACATAGCCCA....................TATGGTTTTTTACCAA 778
 234 eValSerAsnProGlyGlnThrTyrAsnProValAsnGlyProLeuProA 251

 779 CAGGAGGCTCATTTGGCGACAGTGGCTCACCAATGTTTATCTATGATGCC 828
 251 spTyrGlyAlaProGlyAspSerGlySerProLeuPheAlaTyrAspGlu 267

 829 CAAAAGCAAAAGTGGTTAATTAATGGGGTATTGCAAACGGGCAACCCCTA 878
 268 GlnGlnLysLysTrpValIleValAlaValLeuArgAlaTyrAlaGlyIl 284

 879 TATAGGAAAAAGCAATGGCTTCCAGCTGGTTCGTAAAGATTGGTTC.... 924
 284 eAsnGlyAlaThrAsnTrpTrpAsnValIleProThrAspTyrLeuAsnG 301

 925 .....TATGATGAAATCTTTGCTGGAGATACCCATTCAGTATTCTACGAA 969
 301 lnValMetGlnAspAspPheAspAlaProValAspPheValSerGlyLeu 317

 970 CCACGTCAAAATGGGAAATACTCTTTTAACGACGATAATAATGGCACAGG 1019
 318 ProProLeuAsnTrpThrTyr.........AspLysThrSerGlyThrGl 331

1020 AAAAATCAATGCCAAACATGAACACAATTCTCTGCCTAATAGATTAAAAA 1069
 331 y................................................. 331

1070 CACGAACCGTTCAATTGTTTAATGTTTCTTTATCCGAGACAGCAAGAGAA 1119
 332 ..........................ThrLeuSerGlnGlySerLysAsn 339

1120 CCTGTTTATCATGCTGCAGGTGGTGTCAACAGTTATCGACCCAGACTGAA 1169
 340 TrpThrMetHisGlyGlnLysAspAsnAsp..............LeuAs 351

1170 TAATGGAGAAAATATTTCCTTTATTGACGAAGGAAAAGGCGAATTGATAC 1219
 351 nAlaGlyLysAsnLeuValPheSerGlyGln...AsnGlyAlaIleValL 367

1220 TTACCAGCAACATCAATCAAGGTGCTGGAGGATTATATTTCCAAGGAGAT 1269
 367 euLysAspSerValThrGlnGlyAlaGlyTyrLeuGluPheLysAspSer 383

1270 TTTACGGTCTCGCCTGAAAATAACGAAACTTGGCAAGGCGCGGGCGTTCA 1319
 384 TyrThrValSerAlaGluSerGlyLysThrTrpThrGlyAlaGlyIleIl 400

1320 TATCAGTGAAGACAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAAACG 1369
 400 eThrAspLysGlyThrAsnValThrTrpLysValAsnGlyValAlaGlyA 417

1370 ACCGCCTGTCCAAAATCGGCAAAGGCACGCTGCACGTTCAAGCCAAAGGG 1419
 417 spAsnLeuHisLysLeuGlyGluGlyThrLeuThrIleAsnGlyThrGly 433

1420 GAAAACCAAGGCTCGATCAGCGTGGGCGACGGTACAGTCATTTTGGATCA 1469
 434 ValAsnProGlyGlyLeuLysThrGlyAspGlyThrValValLeuAsnGl 450

1470 GCAGGCAGACGATAAAGGCAAAAAACAAGCCTTTAGTGAAATCGGCTTGG 1519
 450 nGlnAlaAspThrAlaGlyAsnValGlnAlaPheSerSerValAsnLeuA 467

1520 TCAGCGGCAGGGGTACGGTGCAACTGAATGCCGATAATCAGTTCAACCCC 1569
 467 laSerGlyArgProThrValValLeuGlyAspAlaArgGlnValAsnPro 483

1570 GACAAACTCTATTTCGGCTTTCGCGGCGGACGTTTGGATTTAAACGGGCA 1619
 484 AspAsnIleSerTrpGlyTyrArgGlyGlyLysLeuAspLeuAsnGlyAs 500

1620 TTCGCTTTCGTTCCACCGTATTCAAAATACCGATGAAGGGGCGATGATTG 1669
 500 nAlaValThrPheThrArgLeuGlnAlaAlaAspTyrGlyAlaValIleT 517

1670 TCAACCACAATCAAGACAAAGAATCCACCGTT.................. 1701
 517 hrAsnAsnAlaGlnGlnLysSerArgLeuLeuLeuAspLeuLysAlaGln 533

1701 .................................................. 1701
 534 AspThrAsnValSerValProIleGlySerIleSerProPheGlyGlyTh 550

1701 .................................................. 1701
 550 rGlyThrProGlyAsnLeuTyrSerMetIleLeuAsnGlyGlnThrArgP 567

1702 ................................ACCATTACAGGCAATAAA 1719
 567 heTyrIleLeuLysSerAlaSerTyrGlyAsnThrLeuTrpGlyAsnSer 583

1720 .....................GATATTGCTACAACCGGCAATAACAACAG 1748
 584 LeuAsnAspProAlaGlnTrpGluPheValGlyThrAspLysAsnLysAl 600

1749 CTTGGAT..............................AGCAAAAAAGAAA 1768
 600 aValGlnThrValLysAspArgIleLeuAlaGlyArgAlaLysGlnProV 617

1769 TTGCCTACAACGGTTGGTTTGGCGAGAAAGATACGACCAAAACGAACGGG 1818
 617 alIlePheHisGly......................GlnLeuThrGly 625

1819 CGGCTCAACCTTGTTTACCAGCCCGCCGCAGAAGACCGCACCCTGCTGCT 1868
 626 AsnMetAspValThrIleProGlnLeuProGlyGlyArgLysValIleLe 642

1869 TTCCGGCGGAACAAATTTA...AACGGCAACATCACGCAAACAAACGGCA 1915
 642 uAspGlySerValAsnLeuProGluGlyThrLeuSerGluAspSerGlyT 659

1916 AACTGTTTTTCAGCGGCAGACCAACACCGCACGCCTACAATCATTTAAAC 1965
 659 hrLeuIlePheGlnGlyHisProValIleHisAla............... 670

1966 GACCATTGGTCGCAAAAAGAGGGCATTCCTCGCGGGGAAATCGTGTGGGA 2015
 671 .........SerValSerGlySerAlaProValSer......LeuAsnGl 682

2016 CAACGACTGGATCAACCGCACATTTAAAGCGGAAAACTTCCAAATTAAAG 2065
 682 nLysAspTrpGluAsnArgGlnPheIleMetLysThrLeuSerLeuLysA 699

2066 GCGGACAGGCGGTGGTTTCCCGCAATGTT.........GCCAAAGTGAAA 2106
```

```
         ::::::      :::||||||||:::              :::   :::|||
 699 spAlaAspPheHisLeuSerArgAsnAlaSerLeuAsnSerAspIleLys 715

2107 GGCGAT...TGGCATTTGAGCAATCACGCCCAAGCAGTTTTT........ 2145
     :::|||       |||:::::       ::::::  ||||||
 716 SerAspAsnSerHisIleThrLeuGlySerAspArgValPheValAspLy 732

2146 .......GGTGTCGCACCGCAT......................CAAA 2164
           |||:::::   :::
 732 sAsnAspGlyThrGlyAsnTyrValIleLeuGluGluGlyThrSerValP 749

2165 GCCACACAATCTGTACACGTTCGGACTGGACGGGT.............. 2199
        |||:::      ||||||::::::  |||
 749 roAspThrValAsnAspArgSerGlnTyrGluGlyAsnIleThrLeuAsp 765

2200 ............................CTGACAAATTGTGTCGAAAA 2219
                                  :::||||::   :::|||
 766 HisAsnSerThrLeuAspIleGlySerArgPheThrGlyGlyIleGluAl 782

2220 A...............ACCATTACCGACGATAAAGTGATTGCTTCATTGA 2254
                     :::||||||:::     |||:::  :::
 782 aTyrAspSerAlaValSerIleThrSerProAspValLeuLeuThrAlaP 799

2255 CTAAGACCGACATCAGCGGCAATGTCGATCTTGCCGATCACGCTCATTTA 2304
         :::      ::::::::::::   :::   |||    :::||||||
 799 roGlyAlaPheAlaGlySerSerLeuThrValHisAspGlyGlyHisLeu 815

2305 AAT...CTCACAGGGCTTGCCACACTCAACGGCAATCTTAGTGCAAATGG 2351
     :::    |||:::||||||   :::     :::|||::::::::|||:::
 816 ThrAlaLeuAsnGlyLeuPheSer...AspGlyHisIleGlnAlaGlyLy 831

2352 CGATACACGTTATACAGTCAGCCACAACGCCACCCAAAAC.......... 2391
      :::::::::    |||:::|||   :::   :::::::::
 831 sAsnSerLysIleThrLeuSerGlyThrProValLysAspThrAlaAsnG 848

2392 ........GGCAACCTTAGCCTCGTGGGCAATGCCCAAGCAACATTTAAT 2433
                     :::   |||:::   :::   :::   |||   :::
 848 lnTyrAlaProAlaValTyrLeuThrAspGlyTyrAspLeuThrGlyAsp 864

2434 CAAGCCACATTAAACGGCAACACATCGGCTTCGGGCAATGCTTCATTTAA 2483
     :::||||||||||:::   :::   :::|||   ::::::::::::::::
 865 AsnAlaThrLeuGluIleThrArgGlyAlaHisAlaSerGlyAspIleHi 881

2484 TCTAAGC...GACCACGCCGTACAAAACGGCAGTCTGACGCTTTCCGGCA 2530
     :   |||         :::|||      ||||||   |||   :::
 881 sAlaSerAlaAlaSerThrValThrIleGlySerAspThrProAlaGluL 898

2531 ACGCTAAGGCAAACGTAAGCCATTCCGCACTCAACGGTAATGTCTCCCTA 2580
       |||:::|||:::::::::      |||||||:::   |||::::::
 898 euAlaSerAlaGluThrThrAlaSerAlaPheAlaGlySerLeuLeuGlu 914

2581 GCCGATAAGGCAGTATTCCATTTTGAAAGCAGCCGCTTTACCGGACAAAT 2630
     :::    :::|||:::|||:::                       |||   ||.
 915 GlyTyrAsnAlaAlaPheAsn.....................GlyAlaIl 924

2631 CAGCGGCGGCAAG...GATACGGCATTACACTTAAAAGACAGCGAATGGA 2677
     |:::||||||:::     |||:::::::::|||        ::::::   ||||
 924 eThrGlyGlyArgAlaAspValSerMetHis......AsnAlaLeuTrpT 939

2678 CGCTGCCGTCAGGCACGGAATTAGGCAATTTAAACCTTGACAACGCCACC 2727
      |||||   :::   :::   :::   :::|||::::::   |||:::
 939 hrLeuGlyGlyAspSerAlaIleHisThrLeuThrValArgAsnSerArg 955

2728 ATTACACTCAATTCCGCCTATCGCCACGATGCGGCAGGGGCGCAAACCGG 2777
      |||:::                              :::   |||   :::|||
 956 IleSer.........................SerGluGlyAspArgThr.. 963

2778 CAGTGCGACAGATGCGCCGCGCCGCCGTTCGCGCCGTTCGCGCCGTTCCC 2827

 963 .................................................. 963

2828 TATTATCCGTTACACCGCCAACTTCGGTAGAATCCCGTTTCAACACGCTG 2877
                                            |||:::||||||

 964 ......................................PheArgThrLeu 967

2878 ACGGTAAACGGCAAATTGAACGGTCAGGGAACA...TTCCGCTTTATGTC 2924
     |||||||||   ||||||::::::   |||:::   |||   :::   ::
 968 ThrValAsn...LysLeuAspAlaThrGlySerAspPheValLeuArgTh 983

2925 GGAACTCTTCGGCTACCGCAGCGACAAATTGAAGCTGGCGGAAAGTTCCG 2974
     :::|||          ::::::||||||::::::::::::|||::::::
 983 rAspLeu......LysAsnAlaAspLysIleAsnValThrGluLysAlaT 998

2975 AAGGCACTTACACCTTGGCGGTCAACAATACCGGCAACGAACCTGCAAGC 3024
      |||:::
 998 hrGlySer.......................................... 1000

3025 CTCGAACAATTGACGGTAGTGGAAGGAAAAGACAACAAACCGCTGTCCGA 3074
                                    ||||||         |||
1001 ..............................AspAsn.........SerLe 1004

3075 AAACCTTAATTTCACCCTGCAAAACGAACACGTCGATGCCGGCGCGTGGC 3124
      |||:::::|||
1004 uAsnValSerPhe..................................... 1008

3125 GTTACCAACTCATCCGCAAAGACGGCGAGTTCCGCCTGCATAATCCGGTC 3174
                                       :::   :::|||:::
1009 ..................................MetLysAspProAla 1013

3175 AAAGAACAAGAGCTTTCCGACAAACTCGGCAAGGCAGAAGCCAAAAAACA 3224
      :::   |||:::|||:::         |||      |||   |||      ::
1014 GlnGlyGlnSerLeuAsnIleProLeuValThrAlaProAlaGlyThrSe 1030

3225 GGCGGAAAAAGACAACGCGCAAAGCCTTGACGCGCTGATTGCGGCCGGGC 3274
      :||||||      :::|||   :::                        |||
1030 rAlaGluMetPheLysAlaGlyThrArg............MetIleGlyP 1043

3275 GCGATGCCGTCGAAAAGACAGAAAGCGTTGCCGAACCGGCCCGGCAGGCA 3324
       :::   |||      |||      |||                  :::
1043 heSerArgValThrProThrLeuHisVal...........AspThrSer 1055

3325 GGCGGGGAAAATGTCGGCATTATGCAGGCGGAGGAAGAGAAAAAACGGGT 3374
     ||||||::::::      |||::::::::::
1056 GlyGlyAsnThrLysTrpIleLeuAspGly..................Ph 1066

3375 GCAGGCGGATAAAGACACCGCCTTGGCGAAACAGCGCGAAGCGGAAACCC 3424
      :::|||:::   |||   |||   |||   :::   ::::::
1066 eLysAlaGluAlaAspLysAlaAlaAlaAlaLysAlaAspSer....... 1080

3425 GGCCGGCTACCACCGCCTTCCCCCGCGCCCGCCGCGCCCGCCGGGATTTG 3474

1080 .................................................. 1080

3475 CCGCAACTGCAACCCCAACCGCAGCCCCAACCGCAGCGCGACCTGATCAG 3524

1080 .................................................. 1080

3525 CCGTTATGCCAATAGCGGTTTGAGTGAATTTTCCGCCACGCTCAACAGCG 3574
         :::   |||:::|||   ::::::|||   :::   :::|||:::
1081 ....PheMetAsnAlaGlyTyrLysAsnPheMetThrGluValAsnAsnL 1096

3575 TTTCGCCGTACAGGACGAATTAGACCGCGTATTTGCCGAAGACCGCCGC 3624
     ::                                          ::::::|||
1096 eu.................................AsnLysArg 1099

3625 AACGCCGTTTGGACAAGCGGCATCCGGGACACCAAACACTACCGTTCGCA 3674
                 :::   :::|||||||||                :::
1100 ...........MetGlyAspLeuArgAspThr...........AsnGl 1108

3675 AGATTTCCGCGCCTACCGCCAACAAACCGACCTGCGCCAAATCGGTATGC 3724
       |||      |||:::   :::                            ::::
1108 yAspAlaGlyAlaTrpAlaArg........................IleM 1117

3725 AGAAAAACCTCGGCAGCGGGCGCGTCGGCATCCTGTTTTCGCACAACCGG 3774
      ::::::::   |||||||:::      |||         ::::::::::::
1117 etSerGlyAlaGlySerAlaAspGlyGlyTyrSerAspAsnTyrThrHis 1133
```

```
3775 ACCGAAAACACCTTCGAC..............GACGGCATCGGC..... 3804
     ::::::       ||||||            |||||||:::
1134 ValGlnValGlyPheAspLysLysHisGluLeuAspGlyValAspLeuPh 1150

3805 ........................AACTCGGCACGGCTTGCCCACGGC. 3828
                         :::|||:::      :::|||.::: 
1150 eThrGlyValThrMetThrTyrThrAspSerSerAlaAspSerHisAlaP 1167

3829 .................GCCGTTTTCGGGCAATACGGCATCGACAGGTTC 3861
                  :::|||   |||
1167 heSerGlyLysThrLysSerValGlyGly.................... 1176

3862 TACATCGGCATCAGCGCGGGCGCGGGTTTAGCAGCGGCAGCCTTTCAGA 3911
              |||:::   |||:::|||   |||:::|||||||:::      ||
1177 ......GlyLeuTyrAlaSerAlaLeuPheGluSerGlyAlaTyrIleAs 1191

3912 CGGCATCGGAGGCAAAATCCGCCGCCGCGTGCTGCATTACGGCATTCAGG 3961
     |   ||||||              :::   :::|||:::      :::
1191 pLeuIleGly...............LysTyrIleHisHis......AspA 1201

3962 CACGATACCGCGCCGGTTTCGGCGGATTCGGCATCGAACCGCACATCGGC 4011
        |||   ::::::|||:::|||:::|||
1201 snAspTyrThrGlyAsnPheAlaGlyLeuGly.................. 1211

4012 GCAACGCGCTATTTCGTCCAAAAAGCGGATTACCGCTACGAAAACGTCAA 4061
       |||:::::::::            :::   |||
1212 ...ThrLysHisTyrAsnThrHisSerTrpTyr................ 1221

4062 TATCGCCACCCCCGGCCTTGCATTCAACCGCTACCGCGCGGGCATTAAGG 4111
                                      ||||||   ::::
1222 ...................................AlaGlyAlaGluT 1226

4112 CAGATTATTCATTCAAACCGGCGCAACACATTTCCATCACGCCTTATTTG 4161
     ::   |||   :::       :::::::::        |||   |||
1226 hrGlyTyrArgTyrHisLeuThrGluGluThrPheIleGluProGlnAla 1242

4162 AGCCTGTCCTATACCGATGCCGCTTCGGGCAAA................ 4194
     :::|||   |||        |||:::|||||||||
1243 GluLeuValTyr...GlyAlaValSerGlyLysThrPheArgTrpLysAs 1258

4195 ..................GTCCGAACACGCGTCAATACCGCCGTATTGG 4225
                   :::::::::|||      :::    ::::::
1258 pGlyAspMetAspLeuSerMetLysAsnArgAspPheSerProLeuIleG 1275

4226 CTCAG.........GATTTCGGCAAAACCCGCAGTGCG...GAATGGGGC 4263
     :::::          ::::::|||||||||   |||:::   :::|||:::
1275 lyArgThrGlyIleGluLeuGlyLysThrPheSerGlyLysAspTrpSer 1291

4264 GTAAACGCC 4272
     |||:::|||
1292 ValThrAla 1294
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:ABB52592

```
seq_documentation_block:
ID   ABB52592 standard; Protein; 1376 AA.
XX
AC   ABB52592;
XX
DT   11-FEB-2002  (first entry)
XX
DE   Escherichia coli polypeptide SEQ ID NO 560.
XX
KW   Escherichia coli; B2/D+A-; antiinflammatory; antibacterial;
KW   immunosuppressive; extra-intestinal infection; phylogeny; meningitis;
KW   systemic infection; non-diarrhoeal infection; septicaemia;
KW   pyelonephritis; antibiotic resistance.
XX
OS   Escherichia coli.
XX
PN   WO200166572-A2.
XX
PD   13-SEP-2001.
XX
PF   12-MAR-2001; 2001WO-EP03445.
XX
PR   10-MAR-2000; 2000FR-0003145.
PR   02-FEB-2001; 2001FR-0001449.
XX
PA   (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
XX
PI   Bingen E,  Bonacorsi S,  Clermont O,  Nassif X,  Tinsley C;
XX
DR   WPI; 2001-550253/61.
XX
PT   A library of DNA fragments of Escherichia coli strains for the
PT   phylogenic determination of a given strain comprises polynucleotides of
PT   nature B2/D+ A- -
XX
PS   Example 6; Fig 6; 646pp; English.
XX
CC   The invention relates to a library of DNA fragments of Escherichia coli
CC   strains comprising polynucleotides (ABA88577-ABA88729 and ABA89533)
CC   and encoded proteins (ABB52459-ABB52919 and ABB52954-ABB53094) of nature
CC   B2/D+A-. The polynucleotides have potential antiinflammatory,
CC   antibacterial and immunosuppressive activity as part of pharmaceutical
CC   compositions used to treat, palliate or prevent extra-intestinal E. coli
CC   infections. The polypeptides are useful for determining the phylogenic
CC   group of a given E. coli strain. These polypeptides can detect and treat
CC   an undesired development of E. coli, particularly an extra-intestinal
CC   infection that include systemic and non-diarrhoeal infections such as
CC   septicaemia, pyelonephritis and meningitis this is particularly
CC   advantageous as bacterial resistance is increasing with the more
CC   frequent use of broad spectrum antibiotics.
XX
SQ   Sequence   1376 AA;
```

```
alignment_scores:
           Quality: 1053.00              Length:   1585
             Ratio:  1.295                 Gaps:     66
 Percent Similarity:  51.293   Percent Identity: 25.363

alignment_block:
 US-09-303-518D-649 x ABB52592   ..

 Align seg 1/1  to: ABB52592  from: 1  to: 1376

  10 ACCGACAAACGGACAACCGAAACACACCGCAAAGCCCCGAAA......AC 53
      |||   |||   :::::::|||      |||::::::   |||      ||
  15 ThrValLysValValSerGluLeuAlaArgArgValCysLysGlySerTh 31

  54 CGGCCGCATCCGCTTCTCGCCTGCTTACTTAGCCATATGCCTGTCGTTCG 103
      |   |||   :::         :::      :::::::::||||||   :
  31 rArgArgGlyLysArgLeuSerValLeuThrSerLeuAlaLeuSer...A 47

 104 GCATTCTTCCCCAAGCCTGGGCGGGACACACTTATTTCGGCATCAAC... 150
      ::::::||||||   :::   ::::::   |||      |||   |||
  47 laLeuLeuProThrValAlaGlyAlaSerThrValGlyGlyAsnAsnPro 63

 151 TACCAATACTATCGCGACTTTGCCGAAAATAAAGGCAAGTTTGCAGTCGG 200
     ||||||   ||||||||||||||||||||||||||||:::|||   :::||
  64 TyrGlnThrTyrArgAspPheAlaGluAsnLysGlyGlnPheGlnAlaGl 80

 201 GGCGAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTGGTCGGCAAAT 250
      ||||   :::|||   ::::::|||:::||||||||||||||||||
  80 yAlaThrAsnIleProIlePheAsnAsnLysGlyGluLeuValGly...H 96

 251 CAATGACAAAAGCCCCGATGATTGATTTTTCT.........GTGGTGTCG 291
        :::   ||||||||||||:::|||||||||         |||   |||
  96 isLeuAspLysAlaProMetValAspPheSerSerValAsnValSerSer 112

 292 CGTAACGGCGTGGCGGCATTGGTGGGCGATCAATATATTGTGAGCGTGGC 341
      :::   |||||||||:::||||::::::   |||||||||:::||||||
 113 AsnProGlyValAlaThrLeuIleAsnProGlnTyrIleAlaSerValLy 129
```

```
 342 ACATAACGGCGGCTATAACAACGTTGATTTTGGTGCGGAAGGAAGAAATC 391
      ||||||    ||||||:::::|||:::||||||    :::|||:::|||
 129 sHisAsnLysGlyTyrGlnSerValSerPheGly...AspGlyGlnAsn. 144

 392 CCGATCAACATCGTTTACTTATAAAATTGTGAAACGGAATAATTATAAA 441
                         :::|||    ||||||   ||||||:::::::::
 145 .................SerTyrHisIleValAspArgAsnGluHisSer 155

 442 GCAGGGACTAAAGGCCATCCTTATGGCGGCGATTATCATATGCCGCGTTT 491
     :::                        :::|||   |||   ||||||||
 156 Ser........................SerAspLeuHisThrProArgLe 164

 492 GCATAAATTTGTCACAGATGCAGAACCTGTTGAAATGACCAGTTATATGG 541
     |    |||:::||||||::::::    |||:::   :::|||||||      :
 164 uAspLysLeuValThrGluValAlaProAlaThrValThrSer...SerS 180

 542 ATGGGCGGAAATATATCGATCAAAATAATTACCCTGACCGTGTTCGTATT 591
     ::               ::::::    ::::::|||              |||
 180 erThrAlaAspIleLeuAsnProSerLysTyrSerAlaPheTyrArgAla 196

 592 GGGGCAGGCAGGCAATATTGGCGATCTGATGAAGATGAGCCCAATAACCG 641
     |||:::|||    ||||||    ::::::::::::    :::
 197 GlySerGlySerGlnTyrIleGlnAspSerGlnGlyLys........... 209

 642 CGAAAGTTCATATCATATTGCAAGTGCGTATTCTTGGCTCGTTGGTGGC. 690
               :::    ::::::::::::|||::::::|||:::||||||
 210 .......ArgHisTrpValThrGlyGlyTyrGlyTyrLeuThrGlyGlyI 224

 691 ........AATACCTTTGCACAAAATGGATCAGGTGGTGGCACAGTCAAC 732
             :::::::|||      :::||||||   |||      :::
 224 leLeuProThrSerPhePheTyrHisGlySerAspGlyIleGlnLeuTyr 240

 733 TTAGGTAGTGAAAAAATTAAACATAGCCCATATGGTTTTTACCAACAGG 782
     :::|||::::::              ||||||          :::|||||||:::
 241 MetGlyGlyAsnIleHisAspHisSer.........IleLeuProSerPh 254

 783 AGGCTCATTTGGCGACAGTGGCTCACCAATGTTTATCTATGATGCCCAAA 832
       |||:::   |||||||||||||||||||:::|||    :::::::::    |
 254 eGlyGluAlaGlyAspSerGlySerProLeuPheGlyTrpAsnThrAlaL 271

 833 AGCAAAAGTGGTTAATTAATGGGGTATTGCAAACGGGCAACCCCTATATA 882
      ||    :::|||    :::    ||||||                     |||
 271 ysGlyGlnTrpGluLeuValGlyVal.................TyrSer 281

 883 GGAAAAAGCAATGGCTTCCAGCTGGTTCGTAAA.........GATTGGTT 923
     |||    ::::::|||    :::|||:::    :::          ::: ,  ||
 282 GlyValGlyGlyGlyThrAsnLeuIleTyrSerLeuIleProGlnSerPh 298

 924 CTATGATGAAATCTTTGCTGGAGATACCCATTCA...GTATTCTACGAAC 970
     |    ::::::|||::::::    |||:::    :::    ||||||::::::
 298 eLeuSerGlnIleTyrSerGluAspAsnAspAlaProValPhePheAsnA 315

 971 CACGTCAAAATGGGAAATACTCTTTTAAC...GACGATAATAATGGCACA 1017
         :::::::::         :::::::::    |||:::::::::||||||
 315 laSerSerGlyAlaProLeuGlnTrpLysPheAspSerSerThrGlyThr 331

1018 GGAAAAATCAATGCCAAACATGAACACAATTCTCTGCCTAATAGATTAAA 1067
     |||
 332 Gly............................................... 332

1068 AACACGAACCGTTCAATTGTTTAATGTTTCTTTATCCGAGACAGCAAGAG 1117
                                 ||||||::::::    :::    |
 333 ...........................SerLeuLysGlnGlySerAspG 340

1118 AACCTGTTTATCATGCTGCAGGTGGTGTCAACAGTTATCGACCCAGACTG 1167
      ||    :::    |||:::       |||    :::              |||
 340 luTyrAlaMetHisGlyGlnLysGlySerAsp...............Leu 351

1168 AATAATGGAGAAAATATTTCCTTTATTGACGAAGGAAAAGGCGAATTGAT 1217
      |||    |||:::|||:::::::|||:::    :::    :::|||::::::
 352 AsnAlaGlyLysAsnLeuThrPheLeuGlyHis...AsnGlyGlnIleAs 367

1218 ACTTACCAGCAACATCAATCAAGGTGCTGGAGGATTATATTTCCAAGGAG 1267
       |||    ::::::::::::|||||||||||||:::|||    |||    |
 367 pLeuGluAsnSerValThrGlnGlyAlaGlySerLeuThrPheThrAspA 384

1268 ATTTACGGTCTCGCCTGAAAATAACGAAACTTGGCAAGGCGCGGGCGTT 1317
     ||:::||||||:::    :::|||::::::||||||    ||||||||||:::
 384 spTyrThrValThrThrSerAsnGlySerThrTrpThrGlyAlaGlyIle 400

1318 CATATCAGTGAAGACAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAAA 1367
       :::::::::|||::::::|||:::|||:::||||||||||||    ::
 401 IleValAspLysAspAlaSerValAsnTrpGlnValAsnGlyValLysGl 417

1368 CGACCGCCTGTCCAAAATCGGCAAAGGCACGCTGCACGTTCAAGCCAAAG 1417
      :|||:::|||    |||||||||:::|||||||||    ||||||:::    |
 417 yAspAsnLeuHisLysIleGlyGluGlyThrLeuValValGlnGlyThrG 434

1418 GGGAAAACCAAGGCTCGATCAGCGTGGGCGACGGTACAGTCATTTTGGAT 1467
      ||    |||:::|||::::::::::|||||||||||||||||:::|||:::
 434 lyValAsnGluGlyGlyLeuLysValGlyAspGlyThrValValLeuAsn 450

1468 CAGCAGGCAGACGATAAAGGCAAAAAACAAGCCTTTAGTGAAATCGGCTT 1517
     |||||||||||:::::::|||         ||||||||||::::::::::::
 451 GlnGlnAlaAspSerSerGlyHisValGlnAlaPheSerSerValAsnIl 467

1518 GGTCAGCGGCAGGGGTACGGTGCAACTGAATGCCGATAATCAGTTCAACC 1567
      ::::|||||||||    ||||||    |||         :::::::|||    ||||
 467 eAlaSerGlyArgProThrValValLeuAlaAspAsnGlnGlnValAsnP 484

1568 CCGACAAACTCTATTTCGGCTTTCGCGGCGGACGTTTGGATTTAAACGGG 1617
      ||||||:::::    :::|||:::|||||||||    ||||||:::|||||||
 484 roAspAsnIleSerTrpGlyTyrArgGlyGlyValLeuAspValAsnGly 500

1618 CATTCGCTTTCGTTCCACCGTATTCAAAATACCGATGAAGGGGCGATGAT 1667
     ::::::|||:::||||||::::::::     :::|||    ||||||    ::
 501 AsnAspLeuThrPheHisLysLeuAsnAlaAlaAspTyrGlyAlaThrLe 517

1668 TGTCAACCACAATCAAGACAAAGAATCCACCGTTACCATT.........A 1708
      :    |||    ::::::||||||    :::::::::|||:::           |
 517 uGlyAsn...SerSerAspLysThrAlaAsnIleThrLeuAspTyrGlnT 533

1709 CAGGCAATAAAGATATTGCT...........ACAACCGGCAATAACAAC 1746
      ||          |||:::              :::::::::|||::::::
 533 hrArgProAlaAspValLysValAsnGluTrpSerSerSerAsnArgGly 549

1747 AGCTTGGATAGCAAAAAAGAAATTGCCTACAAC................ 1779
     ::::::    |||            ||||||
 550 ThrValGlySerLeu......TyrIleTyrAsnAsnProTyrThrHisTh 564

1780 ................................GGTTGGTTTGGCGAGA 1795
                                     |||||||||
 564 rValAspTyrPheIleLeuLysThrSerSerTyrGlyTrpPheProThrG 581

1796 AAGATACGACCAAAACG................................ 1812
         ::::::::::::
 581 lyGlnValSerAsnGluHisTrpGluTyrValGlyHisAspGlnAsnSer 597

1813 ..............................................AA 1814
                                                   ::
 598 AlaGlnAlaLeuLeuAlaAsnArgIleAsnAsnLysGlyTyrLeuTyrHi 614

1815 CGGGCGGCTC......AACCTTGTTTACCAGCCCGCCGCAGAAGACCGCA 1858
      :|||:::|||         |||:::    ::::::          |||        |
 614 sGlyLysLeuLeuGlyAsnIleAsnPheSerAsnLysAlaThrProGlyT 631

1859 CC.........CTGCTGCTTTCCGGCGGAACAAATTTAAACGGCAACATC 1899
      ||             |||:::::::::|||::::::|||::::::|||::::::
 631 hrThrGlyAlaLeuValMetAspGlySerAlaAsnMetSerGlyThrPhe 647

1900 ACGCAAACAAACGGCAAACTGTTTTCAGCGGCAGACCAACACCGCACGC 1949
     ||||||    ||||||:::|||    ::::::|||:::|||:::    |||||
 648 ThrGlnGluAsnGlyArgLeuThrIleGlnGlyHisProValIleHisAl 664

1950 C.........................TACAATCATTTAAACGACCATT 1972
```

```
 664 aSerThrSerGlnSerIleAlaAsnThrValSerSerLeuGlyAspAsn. 680

1973 GGTCGCAAAAAGAGGGCATTCCTCGCGGGGAAATCGTGTGGGACAACGAC 2022
 681 ..........SerValLeuThrGlnProThrSerPheThrGlnAspAsp 693

2023 TGGATCAACCGCACATTTAAAGCGGAAAACTTCCAAATTAAAGGCGGACA 2072
 694 TrpGluAsnArgThrPheSerPheGlySerLeuValLeuLysAspThrAs 710

2073 GGCGGTGGTTTCCCGCAATGTTGCCAAAGTGAAAGGCGATTGGCATTTGA 2122
 710 pPheGlyLeuGlyArgAsnAlaThr....................LeuA 720

2123 GCAATCACGCCCAAGCA...GTTTTGGTGTCGCACCGCATCAAAGCCAC 2169
 720 snThrThrIleGlnAlaAspAsnSerSerValThrLeuGlyAspSerArg 736

2170 ACAATCTGTACACGTTCGGACTGGACGGGTCTGACAAATTGTGTCGAAAA 2219
 737 ValPheIleAspLysLysAspGlyGlnGlyThrAlaPheThrLeuGluGl 753

2220 A.........ACCATTACCGACGATAAAGTGATTGCTTCATTGACTAAGA 2260
 753 uGlyThrSerValAlaThrLysAsp..............AlaAspLysS 765

2261 CCGACATCAGCGGCAATGTCGATCTTGCCGATCACGCTCATTTAAATCTC 2310
 765 erValPheAsnGlyThrValAsnLeuAspAsnGlnSerValLeuAsnIle 781

2311 ACAGGGCTTGCCACACTCAACGGCAATCTTAGTGCAAATGGCGATACACG 2360
 782 AsnGluIle......PheAsnGlyGlyIleGlnAlaAsnAsnSerThrVa 796

2361 TTATACAGTCAGCCACAACGCCACCCAAAACGGCAACCTTAGCCTCGTGG 2410
 796 lAsnIleSerSerAspSerAla....................ValLeuG 806

2411 GCAATGCCCAAGCAACATTTAATCAAGCCACATTAAACGGCAACACATCG 2460
 806 luAsnSerThrLeuThrSerThrAlaLeuAsnLeuAsnLysGlyAlaAsn 822

2461 GCTTCGGGCAATGCTTCATTT..................AATCTAAGCGA 2492
 823 ValLeuAlaSerGlnSerPheValSerAspGlyProValAsnIleSerAs 839

2493 CCACGCCGTACAA..................................... 2505
 839 pAlaThrLeuSerLeuAsnSerArgProAspGluValSerHisThrLeuL 856

2506 .................AACGGCAGTCTGACGCTTTCCGGC...AACGCT 2535
 856 euProValTyrAspTyrAlaGlySerTrpAsnLeuLysGlyAspAspAla 872

2536 AAGGCAAACGTAAGC...CATTCCGCACTCAACGGTAATGTCTCCCTAGC 2582
 873 ArgLeuAsnValGlyProTyrSerMetLeuSerGlyAsnIleAsnValGl 889

2583 CGATAAG........................................... 2589
 889 nAspLysGlyThrValThrLeuGlyGlyGluGlyGluLeuSerProAspL 906

2590 .........................GCAGTATTCCATTTTGAAAGCAGC 2613
 906 euThrLeuGlnAsnGlnMetLeuTyrSerLeuPheAsnGlyTyrArgAsn 922

2614 CGCTTTACCGGACAAATCAGCGGCGGCAAGGATACGGCATTACACTTAAA 2663
 923 ThrTrpSerGlySerLeuAsn...AlaProAspAlaThrValSerMetTh 938

2664 AGACAGCGAATGGACGCTGCCGTCAGGCACGGAATTAGGCAATTTAAACC 2713

 938 rAspThrGlnTrpSerMetAsnGlyAsnSerThrAlaGlyAsnMetLysL 955

2714 TTGACAACGCCACCATTACACTCAATTCCGCCTATCGCCACGATGCGGCA 2763
 955 euAsnArgThrIleValGlyPheAsnGlyGly.................. 965

2764 GGGGCGCAAACCGGCAGTGCGACAGATGCGCCGCGCCGCCGTTCGCGCCG 2813
 965 .................................................. 965

2814 TTCGCGCCGTTCCCTATTATCCGTTACACCGCCAACTTCGGTAGAATCCC 2863
 966 ...........................................ThrSerS 968

2864 GTTTCAACACGCTGACGGTAAACGGCAAATTGAACGGTCAGGGAACATTC 2913
 968 erPheThrThrLeuThrThrAspAsnLeuAspAlaValGlnSerAlaPhe 984

2914 CGCTTTATGTCGGAACTCTTCGGCTACCGCAGCGACAAATTGAAGCTGGC 2963
 985 ValMetArgThrAspLeu......AsnLysAlaAspLysLeuValIleAs 999

2964 GGAAAGTTCCGAAGGCACTTACACCTTGGCGGTCAACAATACCGGCAACG 3013
 999 nLysSerAla........................................ 1002

3014 AACCTGCAAGCCTCGAACAATTGACGGTAGTGGAAGGAAAAGACAACAAA 3063
1003 ..............................ThrGlyHisAspAsn... 1007

3064 CCGCTGTCCGAAAACCTTAATTTCACCCTGCAAAACGAACACGTCGATGC 3113
1008 ......SerIleTrpValAsnPhe......................... 1013

3114 CGGCGCGTGGCGTTACCAACTCATCCGCAAAGACGGCGAGTTCCGCCTGC 3163
1013 .................................................. 1013

3164 ATAATCCGGTCAAAGAACAAGAGCTTTCCGACAAACTCGGCAAGGCAGAA 3213
1013 .................................................. 1013

3214 GCCAAAAAACAGGCGGAAAAAGACAACGCGCAAAGCCTTGACGCGCTGAT 3263
1014 LeuLysLysProSerAspLysAsp.........ThrLeuAspIleProLe 1027

3264 TGCGGCCGGGCGCGATGCCGTCGAAAAGACAGAAAGCGTTGCCGAACCGG 3313
1027 u.................ValSerAlaProGluAlaThrAlaAspAsnL 1038

3314 CCCGGCAGGCAGGCGGGGAAAATGTCGGCATT.................. 3345
1038 euPheArgAlaSerThrArgValValGlyPheSerAspValThrProThr 1054

3346 ......ATGCAGGCGGAGGAAGAGAAAAAACGGGTGCAGGCGGATAAAGA 3389
1055 LeuSerValArgLysGluAspGlyLysLysGluTrpValLeuAsp..... 1069

3390 CACCGCCTTGGCGAAACAGCGCGAAGCGGAAACCCGGCCGGCTACCACCG 3439
1069 .................................................. 1069

3440 CCTTCCCCCGCGCCCGCCGCGCCCCGCCGGGATTTGCCGCAACTGCAACCC 3489
1070 ........GlyTyrGlnValAlaArgAsnAsp.................. 1077

3490 CAACCGCAGCCCCAACCGCAGCGCGACCTGATCAGCCGTTATGCCAATAG 3539
1078 ...............GlyGlnGlyLysAlaAlaAlaThrPheMetHisIl 1089

3540 CGGTTTGAGTGAATTTTCCGCCACGCTCAACAGCGTTTTCGCCGTACAGG 3589
1089 eSerTyrAsnAsnPheIleThrGluValAsnAsnLeuAsnLysArgMetG 1106
```

```
3590 ACGAATTAGACCGCGTATTTGCCGAAGACCGCCGCAACGCCGTTTGGACA 3639
       :::|||     :::   :::|||             ::::::|||:::
1106 lyAspLeuArgAspIleAsnGlyGlu.........AlaGlyThrTrpVal 1119

3640 ...............AGCGGCATCCGGGAC.........ACCAAACACTA 3665
                    ||||||      |||         |||   |||||
1120 ArgLeuLeuAsnGlySerGlySerAlaAspGlyGlyPheThrAspHisTy 1136

3666 CCGTTCGCAAGATTTCCGCGCCTACCGCCAACAAACCGACCTGCGCCAAA 3715
     |                                         |||   |:|:
1136 r.....................................ThrLeuLeuGlnM 1141

3716 TCGGTATGCAGAAA......AACCTCGGCAGCGGGCGCGTC......GGC 3753
     ::|||   ::::::      :::|||||||||     :::       |||
1141 etGlyAlaAspArgLysHisGluLeuGlySerMetAspLeuPheThrGly 1157

3754 ATCCTGTTTTCGCACAACCGGACCGAAAACACCTTCGACGACGGCATCGG 3803
     ::::::   :::::::::   |||:::
1158 ValMetAlaThrTyrThrAspThrAsp....................... 1166

3804 CAACTCGGCACGGCTTGCCCACGGCGCCGTTTTCGGGCAATACGGCATCG 3853
         ||||||   |||      |||   :::    :::      |||.
1167 .AlaSerAlaGlyLeuTyrSerGlyLysThrLysSerTrpGlyGly.... 1181

3854 ACAGGTTCTACATCGGCATCAGCGCGGGCGCGGGTTTTAGCAGCGGCAGC 3903
                    |||:::   |||::::::   |||   ||||||:::
1182 ..............GlyPheTyrAlaSerGlyLeuPheArgSerGlyAla 1193

3904 CTTTCAGACGGCATCGGAGGCAAAATCCGCCGCCGCGTGCTGCATTACGG 3953
         |||    |||:::        |||::::::
1194 TyrPheAspLeuIleAlaLysTyrIleHisAsn................. 1204

3954 CATTCAGGCACGATACCGCGCCGGTTTCGGCGGATTCGGCATCGAACCGC 4003
         :::   :::|||       :::|||:::|||   |||    :::
1205 ....GluAsnLysTyrAspLeuAsnPheAlaGlyAlaGlyLysGlnAsnP 1220

4004 ACATCGGCGCAACGCGCTATTTCGTCCAAAAAGCGGATTACCGCTACGAA 4053
         :::   :::   |||            ::::::   |||||||||:::
1220 heArgSerHisSerLeuTyrAlaGlyAlaGluValGlyTyrArgTyrHis 1236

4054 ......AACGTCAATATCGCCACCCCC....................... 4074
         ::::::::::::::   |||
1237 LeuThrAspThrThrPheValGluProGlnAlaGluLeuValTrpGlyAr 1253

4075 .......GGCCTTGCATTCAACCGCTACCGCGCGGGCATTAAGGCAGATT 4117
           |||   :::||||||            :::|||:::       |||
1253 gLeuGlnGlyGlnThrPheAsnTrpAsnAspSerGlyMet......AspV 1268

4118 ATTCATTCAAACCGGCGCAACACATTTCCATCACGCCTTATTTGAGCCTG 4167
      |||::::::         :::   |||:::::::|||         :::
1268 alSerMetArg.........ArgAsnSerValAsnProLeu......Val 1279

4168 TCCTATACCGATGCCGCTTCGGGCAAAGTCCGAACACGCGTCAATACCGC 4217
     :::   |||   :::::::|||||||||:::   :::      :::   ::
1280 GlyArgThrGlyValValSerGlyLysThrPheSerGlyLysAspTrpSe 1296

4218 CGTATTGGCTCAG..............GATTTCGGCAAAACCCGCAGTG 4252
     ::::   |||:::               :::|||       |||   ||||
1296 rLeuThrAlaArgAlaGlyLeuHisTyrGluPheAspLeuThrAspSerA 1313

4253 CGGAA 4257
     ||:::
1313 laAsp 1314
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:AAW27705

seq_documentation_block:

ID   AAW27705 standard; Protein; 323 AA.
XX
AC   AAW27705;
XX
DT   08-MAY-1998 (first entry)
XX
DE   H. influenzae Hap protein autotransporter membrane integration region.
XX
KW   Hap protein; autotransporter; Gram-negative bacteria; diagnostic;
KW   therapy; surface presented polypeptide.
XX
OS   Haemophilus influenzae.
XX
PN   WO9735022-A1.
XX
PD   25-SEP-1997.
XX
PF   15-MAR-1996;   96WO-EP01130.
XX
PR   15-MAR-1996;   96WO-EP01130.
XX
PA   (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
XX
PI   Jose J, Maurer J, Meyer TF;
XX
DR   WPI; 1997-480227/44.
DR   N-PSDB; AAT88142.
XX
PT   Presentation of peptide(s) on surface of Gram-negative bacteria -
PT   via transformation with vector encoding signal peptide, presented
PT   peptide and transporter domain of auto-transporter, producing
PT   peptide libraries for epitope mapping
XX
PS   Claim 8; Fig 9; 84pp; German.
XX
CC   This sequence represents the H. influenzae Hap autotransporter membrane
CC   integration region. This region is involved in a novel method which
CC   allows the presentation of stable fusion polypeptides on the surface of
CC   Gram-negative bacteria which can be released into the surrounding media.
CC   The method can be used to produce a variegated population of
CC   surface-presented polypeptides, so that bacteria expressing polypeptides
CC   with particular properties can be identified and simultaneously selected,
CC   e.g. for epitope mapping or selection of ligands with the highest
CC   affinity for antibodies, major histocompatibility complex (MHC) molecules
CC   or other components of the immune system. Selected polypeptides can be
CC   used diagnostically, e.g. to screen sera or antibody banks, and
CC   polypeptide expressing cells may be used as live vaccines. They may also
CC   be used therapeutically, e.g. when the polypeptide is an antibody, to
CC   remove or concentrate pollutants, inactivate toxins, prepare and process
CC   food, prepare washing compositions and label cells. Selected bacteria can
CC   be stored, reproduced and replicated on a large scale as individual
CC   clones.
XX
SQ   Sequence 323 AA;

alignment_scores:

Quality:	715.50	Length:	307
Ratio:	2.909	Gaps:	1
Percent Similarity:	80.130	Percent Identity:	43.974

alignment_block:
US-09-303-518D-649 x AAW27705 ..

Align seg 1/1 to: AAW27705 from: 1 to: 323

```
3454 CGCCGCGCCCGCCGGGATTTGCCGCAACTGCAACCCCAACCGCAGCCCCA 3503
     ::::::::::::::::|||               :::        :::   ::
  17 GlnGlnSerGluLysAspArgLeuAlaGlnGluGluAlaGluLysGlnAr 33

3504 ACCGCAGCGCGACCTGATCAGCCGTTATGCCAATAGCGGTTTGAGTGAAT 3553
     :   |||:::|||||||||||||||||||:::||||||:::|||||||||:
  33 gLysGlnLysAspLeuIleSerArgTyrSerAsnSerAlaLeuSerGluL 50

3554 TTTCCGCCACGCTCAACAGCGTTTTCGCCGTACAGGACGAATTAGACCGC 3603
     ::||||||||:::||||||:::::::::||||||||||||||||||
  50 euSerAlaThrValAsnSerMetLeuSerValGlnAspGluLeuAspArg 66
```

```
3604 GTATTTGCCGAAGACCGCCGCAACGCCGTTTGGACAAGCGGCATCCGGGA 3653
     :::|||:::::::::  :::::||||||||||||:::       :::||
  67 LeuPheValAspGlnAlaGlnSerAlaValTrpThrAsnIleAlaGlnAs 83

3654 CACCAAACACTACCGTTCGCAAGATTTCCGCGCCTACCGCCAACAA...A 3700
     |   :::::::|||   |||:::  |||||||||||:::||||||   |
  83 pLysArgArgTyrAspSerAspAlaPheArgAlaTyrGlnGlnGlnLysT 100

3701 CCGACCTGCGCCAAATCGGTATGCAGAAAAACCTCGGCAGCGGGCGCGTC 3750
     ||:::|||||||||||||||:::||||||    |||::::::||||||:::
 100 hrAsnLeuArgGlnIleGlyValGlnLysAlaLeuAlaAsnGlyArgIle 116

3751 GGCATCCTGTTTTCGCACAACCGGACCGAAAACACCTTCGACGACGGCAT 3800
     |||    :::||||||||||:::|||::::::||||||||||||:::    ::
 117 GlyAlaValPheSerHisSerArgSerAspAsnThrPheAspGluGlnVa 133

3801 CGGCAACTCGGCACGGCTTGCCCACGGCGCCGTTTTCGGGCAATACGGCA 3850
     :   |||   |||   |||:::      :::   |||:::||||||
 133 lLysAsnHisAlaThrLeuThrMetMetSerGlyPheAlaGlnTyrGlnT 150

3851 TCGACAGGTTCTACATCGGCATCAGCGCGGGCGCGGGTTTTAGCAGCGGC 3900
              :::   :::|||::::::::|||:::|||:::|||::::::
 150 rpGlyAspLeuGlnPheGlyValAsnValGlyThrGlyIleSerAlaSer 166

3901 AGCCTTTCAGACGGCATCGGAGGCAAAATCCGCCGCCGCGTGCTGCATTA 3950
     ::::::::::::        :::     ||||||:::|||:::::::::::::||
 167 LysMetAlaGluGluGlnSerArgLysIleHisArgLysAlaIleAsnTy 183

3951 CGGCATTCAGGCACGATACCGCGCCGGTTTCGGCGGATTCGGCATCGAAC 4000
     ||||:::::::|||   |||:::         :::|||   :::|||||||:::|
 183 rGlyValAsnAlaSerTyrGlnPheArgLeuGlyGlnLeuGlyIleGlnP 200

4001 CGCACATCGGCGCAACGCGCTATTTCGTCCAAAAAGCGGATTACCGCTAC 4050
     ||::::::|||::::::|||||||||||:::::::::     :::|||:::
 200 roTyrPheGlyValAsnArgTyrPheIleGluArgGluAsnTyrGlnSer 216

4051 GAAAACGTCAATATCGCCACCCCCGGCCTTGCATTCAACCGCTACCGCGC 4100
     |||:::|||::::::   ||||||:::||||||||||||||||||||:::||
 217 GluGluValArgValLysThrProSerLeuAlaPheAsnArgTyrAsnAl 233

4101 GGGCATTAAGGCAGATTATTCATTCAAACCGGCGCAACACATTTCCATCA 4150
     ||||||::::::||||||:::|||   |||::::::::::||||||:::
 233 aGlyIleArgValAspTyrThrPheThrProThrAspAsnIleSerValL 250

4151 CGCCTTATTTGAGCCTGTCCTATACCGATGCCGCTTCGGCAAAGTCCGA 4200
       ||||||:::   ::::::|||:::|||::::::::::::::::|||:::
 250 ysProTyrPhePheValAsnTyrValAspValSerAsnAlaAsnValGln 266

4201 ACACGCGTCAATACCGCCGTATTGGCTCAGGATTTCGGCAAAACCCGCAG 4250
     |||    ||||||    :::||||||    |||    ||||||:::        ::
 267 ThrThrValAsnLeuThrValLeuGlnGlnProPheGlyArgTyrTrpGl 283

4251 TGCGGAATGGGGCGTAAACGCCGAAATCAAAGGTTTCACGCTGTCCCTCC 4300
     :   |||    |||::::::|||||||||         |||    :::|||
 283 nLysGluValGlyLeuLysAlaGluIleLeuHisPheGlnIleSerAlaP 300

4301 ACGCTGCCGCCGCCAAAGGCCCGCAACTGGAAGCGCAACACAGCGCGGGC 4350
         :::    ::::::|||    ||||||      |||:::::::::|||
 300 heIleSerLysSerGlnGlySerGlnLeuGlyLysGlnGlnAsnValGly 316

4351 ATCAAATTAGGCTACCGCTGG 4371
     :::||||||||||||||||||
 317 ValLysLeuGlyTyrArgTrp 323
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:AAP80136

seq_documentation_block:

```
ID   AAP80136 standard; protein;  741 AA.
XX
AC   AAP80136;
XX
DT   09-OCT-1990  (first entry)
XX
DE   Neisseria IgA-Protease precursor protein.
XX
KW   IgA-Protease precursor; Neisseria sp;
KW   Gram-negative bacterial live vaccines.
XX
OS   Neisseria gonorrhoeae (strain SMH).
XX
FH   Key             Location/Qualifiers
FT   Cleavage-site   193..197
FT                   /label=IgA-Protease cleavage site (a)
FT   Cleavage-site   226..229
FT                   /label=IgA-Protease cleavage site (b)
FT   Cleavage-site   327..332
FT                   /label=IgA-Protease cleavage site (c)
XX
PN   DE3622221-A.
XX
PD   14-JAN-1988.
XX
PF   02-JUL-1986;   86DE-3622221.
XX
PR   02-JUL-1986;   86DE-3622221.
XX
PA   (PLAC ) MAX PLANCK GES WISSENSCH.
XX
PI   Meyer TF,  Halter R,  Pohlner J;
XX
DR   WPI; 1988-015104/03.
DR   N-PSDB; AAN80154.
XX
PT   Extracellular prodn. of proteins -
PT   by gram-negative host cells contg. a vector contg. one or more
PT   genes coding the desired protein
XX
PS   Disclosure; ; ; German.
XX
CC   Precursor protein consists of three regions i.e. Amino terminal
CC   leader sequence, IgA protease and a "helper" domain. The cleavage
CC   sites given in the features lie in the region between the latter
CC   two domains. DNA encoding a desired protein can be cloned into the
CC   corresponding region in the IgA protease precursor gene, between
CC   the DNA that encodes the natural cleavage sites. Thus, IgA protease
CC   coding region is not disrupted and the desired protein is released
CC   following cleavage by the protease.
XX
SQ   Sequence   741 AA;
```

alignment_scores:

```
           Quality:  420.00            Length:    789
             Ratio:  1.154               Gaps:     19
Percent Similarity:  46.134   Percent Identity: 21.800
```

alignment_block:
US-09-303-518D-649 x AAP80136 ..

Align seg 1/1 to: AAP80136 from: 1 to: 741

```
2521 CTTTCCGGCAACGCTAAGGCAAACGTAAGCCATTCCGCACTCAACGGTAA 2570
     ||||||   :::|||       :::   :::   :::   :::||||||||
   1 LeuSerAspLysAlaLeuAsnSerPheAspAlaThrProIleAsnGlyAs 17

2571 TGTCTCCCTAGCCGATAAGGCAGTATTCCATTTGAAAGCAGCCGCTTTA 2620
     ||||:::|||   ::::::|||::::::   :::   ::::::   :::
  17 nValAsnLeuAsnGlnAsnAlaAlaLeuValLeuGlyLysAlaAlaLeuT 34

2621 CCGGACAAATCAGCGGCGGCAAGGATACGGCATTACACTTA...AAAGAC 2667
        |||:::|||:::|||       ::::::   :::   |||   :::
  34 rpGlyLysIleGlnGlyGlnGlyAsnSerArgValSerLeuAsnGlnHis 50

2668 AGCGAATGGACGCTGCCGTCAGGCACGGAATTAGGCAATTTAAACCTTGA 2717
     |||:::|||   |||   :::   :::::::::   ||||||:::|||
  51 SerLysTrpHisLeuThrGlyAspSerGlnValHisAsnLeuSerLeuAl 67
```

```
2718 CAACGCCACCATTACACTCAATTCCGCCTATCGCCACGATGCGGCAGGGG 2767
       ::::::   |||   ||||||:::|||              ||||||   :::|
  67 aAspSerHisIleHisLeuAsnAsnAla......SerAspAlaGlnSerA 82

2768 CGCAAACCGGCAGTGCGACAGATGCGCCGCGCCGCCGTTCGCGCCGTTCG 2817
     ||
  82 la................................................ 82

2818 CGCCGTTCCCTATTATCCGTTACACCGCCAACTTCGGTAGAATCCCGTTT 2867
                                              ::::::::
  83 ..........................................AsnLysTy 85

2868 CAACACGCTGACGGTAAACGGCAAATTGAACGGTCAGGGAACATTCCGCT 2917
     ::::|||:::    :::|||        |||:::|||:::|||   |||:::
  85 rHisThrIleLysIleAsn...HisLeuSerGlyAsnGlyHisPheHisT 101

2918 TTATGTCGGAACTCTTCGGCTACCGCAGCGACAAATTGAAGCTGGCGGAA 2967
     ::::::::::::|||              :::|||||||:::   :::   |||
 101 yrLeuThrAspLeuAlaLysAsnLeuGlyAspLysValLeuValLysGlu 117

2968 AGTTCCGAAGGCACTTACACCTTGGCGGTCAACAATACCGGCAACGAACC 3017
     |||::::::|||   |||   |||   |||::::|||      ::::||||||
 118 SerAlaSerGlyHisTyrGlnLeuHisValGlnAsnLysThrGlyGluPr 134

3018 TGCAAGCCTCGAACAATTGACGGTAGTGGAAGGAAAAGACAACAAACCGC 3067
     |   :::   |||   |||   :::   :::::::::::   :::
 134 o...AsnGlnGluGlyLeuAspLeuPheAspAlaSerSerValGlnAspA 150

3068 TGTCCGAAAACCTTAATTTCACCCTGCAAAACGAACACGTCGATGCCGGC 3117
       |||   :::|||      :::|||   |||::::::||||||   |||
 150 rgSer...ArgLeuPheValSerLeuAlaAsnHisTyrValAspLeuGly 165

3118 GCGTGGCGTTACCAACTCATCCGCAAAGACGGCGAGTTCCGCCTGCATAA 3167
     |||   ||||||   :::       :::::|||        ||||||:::||
 166 AlaLeuArgTyrThrIleLysThrGluAsnGlyIleThrArgLeuTyrAs 182

3168 T.....................CCGGTCAAAGAACAAGAGCTTTCCGACA 3196
     |                      |||||||||
 182 nProTyrAlaGlyAsnGlyArgProValLysProAlaProSerProAlaA 199

3197 AACTCGGCAAGGCAGAAGCCAAAAAACAGGCGGAAAAAGACAACGCGCAA 3246
                   ::::::|||:::|||   ::::::   |||:::||||||
 199 laAsnThrAlaSerGlnAlaGlnLysAlaThrGlnThrAspGlyAlaGln 215

3247 AGCCTTGACGCGCTGATTGCGGCCGGGCGCGATGCCGTCGAAAAGACAGA 3296
                       ||||||   ::::::   |||
 216 ...............IleAlaLysProGlnAsnIleVal........... 223

3297 AAGCGTTGCCGAACCGGCC............................... 3315
           ||||||   |||:::
 224 ....ValAlaProProSerProGlnAlaAsnGlnAlaGluGluAlaLeuA 239

3316 ..CGGCAGGCAGGCGGGGAAAATGTCGGCATTATGCAGGCG......... 3354
       :::||||||   :::|||:::|||        :::||||||
 239 rgGlnGlnAlaLysAlaGluGlnValLysArgGlnGlnAlaAlaGluAla 255

3355 ......................GAGGAAGAGAAAAAACGGGTGCAGGCGGA 3383
                            :::||||||   ||||||     |||::
 256 GluLysValAlaArgGlnLysAspGluGluAlaLysArgLysAlaAlaGl 272

3384 T................................................. 3384
     :
 272 uIleAlaArgGlnGlnGluGluAlaArgLysAlaAlaGluLeuAlaAlaL 289

3385 .......................AAAGACACCGCCTTGGCGAAACAGCGC 3411
                             |||        ||||||:::|||:::
 289 ysGlnLysAlaGluAlaGluArgLysAlaArgGluLeuAlaArgGlnLys 305

3412 GAAGCGGAAACCCGGCCGGCTACCACCGCCTTCCCCCGCGCCCGCCGCGC 3461
               |||:::          ::::::|||   |||:::   ||||||
 306 AlaGluGluAlaSerHisGlnAlaAsnAlaLysProLysArgArgArgAr 322

3462 CCGCCGGGATTTGCCGCAA............................... 3480
      |||      ||||||:::
 322 gArgAlaIleLeuProArgProProAlaProValPheSerLeuAspAspT 339

3480 ................................................. 3480

 339 yrAspAlaLysAspAsnSerGluSerSerIleGlyAsnLeuAlaArgVal 355

3480 ................................................. 3480

 356 IleProArgMetGlyArgGluLeuIleAsnAspTyrGluGluIleProLe 372

3480 ................................................. 3480

 372 uGluGluLeuGluAspGluAlaGluGluGluArgArgGlnAlaThrGlnP 389

3480 ................................................. 3480

 389 heHisSerLysSerArgAsnArgArgAlaIleSerSerGluProSerSer 405

3480 ................................................. 3480

 406 AspGluAspAlaSerGluSerValSerThrSerAspLysHisProGlnAs 422

3480 ................................................. 3480

 422 pAsnThrGluLeuHisGluLysValGluThrAlaGlyLeuGlnProArgA 439

3481 ..CTGCAACCCCAACCGCAGCCCCAACCGCAGCGCGACCTGATCAGCCGT 3528
          ||||||:::   |||        |||   |||   :::|||
 439 laAlaGlnProArgThrGlnAlaAlaAlaGlnAlaAspAlaValSerThr 455

3529 TATGCCAATAGCGGTTTGAGTGAATTTTCCGCCACGCTCAACAGCGTT.. 3576
        :::||||||:::||||||:::      |||:::   :::|||:::
 456 AsnThrAsnSerAlaLeuSerAspAlaMetAlaSerThrGlnSerIleLe 472

3577 ................TTCGCCGTACAGGACGAATTAGACCGCGTATTTG 3610
                      :::   ::::::         ::::::      |
 472 uLeuAspThrGlyAlaTyrLeuThrArgHisIleAlaGlnLysSerArgA 489

3611 CCGAAGACCGCCGCAACGCCGTTTGGACAAGCGGCATCCGGGACACCAAA 3660
      ||:::   ::::::|||:::||||||   |||:::          :::
 489 laAspAlaGluLysAsnSerValTrpMetSerAsnThrGlyTyrGlyArg 505

3661 CACTACCGTTCGCAAGATTTCCGCGCCTACCGCCAACAAACCGACCTGCG 3710
        |||   |||   ::::::|||   :::   ::::::
 506 AspTyrAlaSerAlaGlnTyrArgArgPheSerSerLysArgThrGlnTh 522

3711 CCAAATCGGTATGCAGAAAAACCTCGGCAGCGGG...CGCGTCGGCATCC 3757
       ||||||||::::::::::::|||:::::::::     ::::::|||   :
 522 rGlnIleGlyIleAspArgSerLeuSerGluAsnMetGlnIleGlyGlyV 539

3758 TGTTTTCGCACAACCGGACCGAAAACACCTTCGACGACGGCATCGGCAAC 3807
     ::::::::::::::   :::::::::||||||||||::::::   |||:::
 539 alLeuThrTyrSerAspSerGlnHisThrPheAspGlnAla...GlyGly 554

3808 TCGGCACGGCTTGCCCACGGCGCCGTTTTCGGGCAATACGGCATC...GA 3854
     :::       ::::::::::::   :::::::|||:::|||   :::   ||
 555 LysAsnThrPheValGlnAlaAsnLeuTyrGlyLysTyrTyrLeuAsnAs 571

3855 CAGGTTCTACATCGGCATCAGCGCGGGCGCGGGTTTAGCAGCGGCAGCC 3904
     |   :::|||::::::   :::   |||||||||         :::   |
 571 pAlaTrpTyrValAlaGlyAspIleGlyAlaGlySerLeuArgSerArgL 588

3905 TTTCAGACGGCATCGGAGGCAAAATCCGCCGCCGCGTGCTGCATTACGGC 3954
     ||:::               ::::::::::::|||   ::::::   |||
 588 euGlnThrGlnGlnLysAlaAsnPheAsnArgThrSerIleGlnThrGly 604

3955 ATTCAGGCACGATACCGCGCCGGTTTCGGCGGATTCGGCATCGAACCGCA 4004
     :::                     ::::::   |||   |||   |||
 605 LeuThrLeuGlyAsnThrLeuLysIleAsnGlnPheGluIleValProSe 621

4005 CATCGGCGCAACGCGCTATTTCGTCCAAAAAGCGGATTACCGCTACGAAA 4054
```

```
        |||                       ::::::::|||||||||:::        :
 621 rAlaGlyIleArgTyrSerArgLeuSerSerAlaAspTyrLysLeuGlyA 638

4055 ACGTCAATATCGCCACCCCCGGCCTTGCATTCAACCGCTACCGCGCGGGC 4104
     ::  ::::::  :::  ::::::|||   :::         ||||||
 638 spAspSerValLysValSerSerMetAlaValLysThrLeuThrAlaGly 654

4105 ATTAAGGCAGATTATTCATTCAAACCGGCGCAACACATTTCCATCACGCC 4154
     :::          |||  ||||||  :::  :::::::::::  ||
 655 LeuAspPheAlaTyrArgPheLys...ValGlyAsnLeuThrValLysPr 670

4155 TTATTTGAGCCTGTCCTATACCGATGCCGCTTCGGGCAAAGTCCGAACAC 4204
     |  ||||||  :::|||     |||     ||||||       ::::
 670 oLeuLeuSerAlaAlaTyrPhe...AlaAsnTyrGlyLysGlyGlyValA 686

4205 GCGTCAATACCGCCGTATTGGCTCAGGATTTCGGCAAAACCCGCAGTGCG 4254
     ::|||:::         :::|||          :::  ::::::
 686 snValGlyGlyLysSerPheAlaTyrLysAlaAspAsnGlnGlnGlnTyr 702

4255 GAATGGGGCGTAAACGCCGAAATCAAAGGTTTCACGCTGTCCCTCCACGC 4304
     :::   ||||||          ::::::   |||||:::::::::
 703 SerAlaGlyValAlaLeuLeuTyrArgAsnValThrLeuAsnValAsnGl 719

4305 TGCCGCCGCCAAAGGCCCGCAACTGGAAGCGCAACACAGCGCGGGCATCA 4354
     ::::   :::||||||  |||||||||  |||   |||:::  ||||
 719 ySerIleThrLysGlyLysGlnLeuGluLysGlnLysSerGlyGlnIleL 736

4355 AATTAGGCTACCGCTGG 4371
     ||:::      |||:::
 736 ysIleGlnIleArgPhe 741
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAG98842

seq_documentation_block:

```
ID   AAG98842 standard; Protein; 1569 AA.
XX
AC   AAG98842;
XX
DT   26-SEP-2001  (first entry)
XX
DE   E. coli growth and proliferation related protein sequence SEQ ID NO:312.
XX
KW   Escherichia coli; growth; proliferation; microbial; antimicrobial;
KW   bacterial infection; microorganism.
XX
OS   Escherichia coli.
XX
PN   WO200134810-A2.
XX
PD   17-MAY-2001.
XX
PF   09-NOV-2000; 2000WO-US30950.
XX
PR   09-NOV-1999;   99US-0164415.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Forsyth RA,  Ohlsen K,  Zyskind J;
XX
DR   WPI; 2001-335933/35.
DR   N-PSDB; AAH84513.
XX
PT   Novel nucleic acids that inhibit Escherichia coli proliferation, useful
PT   for screening for homologous genes and for designing expression vectors
PT   -
XX
PS   Claim 19; Page 396-399; 522pp; English.
XX
CC   AAH84373 to AAH84499 represent Escherichia coli growth and proliferation
CC   related DNA sequences (I). AAH84500 to AAH84670 encode the E. coli
CC   growth and proliferation related proteins given in AAG99078 and AAG98830
CC   to AAG98999. (I) can be used as potential targets for the generation of
CC   new antimicrobial agents, and for identification of compounds which
CC   interact with the gene products of (I). In addition the expression of
CC   (I) and the purification of the proteins, the purified proteins can be
CC   used to generate reagents and screen small molecule libraries or other
CC   candidate compound libraries for compounds that can be further developed
CC   to yield novel antimicrobial compounds. In addition, nucleic acid probes
CC   complementary to (I) that are specific for particular species of
CC   microorganisms can be used to identify particular microorganism species
CC   in clinical specimens, therefore, providing a rapid and dependable
CC   method by which to identify the causative agents of a bacterial
CC   infection. Also, antibodies generated against proteins translated from
CC   mRNA transcribed from proliferation-required sequences can also be used
CC   to screen for specific microorganisms that produce such proteins in a
CC   species-specific manner. AAH84371 and AAH84670 represent sequencing
CC   primers used in the isolation of E. coli growth and proliferation
CC   related sequence, which are used in an example from the present
CC   invention.
XX
SQ   Sequence   1569 AA;
```

alignment_scores:

```
           Quality:  364.50            Length:   1634
             Ratio:  0.485               Gaps:     81
Percent Similarity:  46.022   Percent Identity: 19.951
```

alignment_block:

```
US-09-303-518D-649 x AAG98842   ..

Align seg 1/1  to: AAG98842  from: 1  to: 1569

 445 GGGACTAAAGGCCATCCTTATGGCGGCGATTATCATATGCCGCGTTTGCA 494
     ||||||  ::::::  :::       ::::::  :::   ::::::||
  97 GlyThrThrAsnAsnThrHisIleIleAsnHisGlyGlnGlnAsnValHi 113

 495 TAAATTTGTCACAGATGCAGAACCTGTTGAAATGACCAGTTATATGGAT. 543
     |      |||::::::::::   :::|||      :::|||:::|||
 113 sGlyGlyValSerAsnGlySerLeuIleGluSerGlyGlyTyrGlnAspI 130

 544 ........GGGCGGAAATATATCGATCAAAATAATTACCCTGACCGTGTT 585
             ::::::::::::  |||   |||          ::::::
 130 leGlySerHisAsnAsnPheValGlyGlnAlaAsn.........AsnThr 143

 586 CGTATTGGGGCAGGCAGGCAATATTGGCGATCTGATGAAGATGAGCCCAA 635
        |||::::::|||||||||
 144 ThrIleAsnGlyGlyArgGln............................ 150

 636 TAACCGCGAAAGTTCATATCATATTGCA......AGTGCGTATTCTTGGC 679
                  |||    |||   :::     |||:::  :::   :
 151 .............SerIleHisAspGlyGlyIleSerThrGlyThrThrI 163

 680 TCGTTGGTGGCAATACCTTTGCACAAAATGGA......TCAGGTGGTGGC 723
     ::    :::||||||     :::    :::|||       |||:::|||
 163 leGluSerGlyAsnGlnAspValTyrLysGlyGlyIleSerAsnGlyThr 179

 724 ACAGTCAACTTAGGTAGTGAAAAAAATTAAA...CATAGCCCATATGGTTT 770
      |||::::::   |||::::::::::::::        |||      |||::
 180 ThrIleLysGlyGlyAlaSerArgValGluGlyGlySerAlaAsnGlyIl 196

 771 TTTACCAACAGGAGGCTCA.............................. 789
      :|||     |||||||||
 196 eLeuIleAspGlyGlySerGlnIleValLysValGlnGlyHisAlaAspG 213

 790 ........TTTGGCGACAGTGGCTCACCAATGTTTATCTATGATGCCCAA 831
             ::::::  |||||||||          :::       :::
 213 lyThrThrIleAsnLysSerGlySerGlnAspValValGlnGlySerLeu 229

 832 AAGCAAAAGTGGTTAATTAATGGGGTATTGCAAACGGGCAACCCCTATAT 881
          :::         |||||||||             |||:::  |||::
 230 AlaThrAsnThrThrIleAsnGly............GlyArgGlnTyrVa 242

 882 AGGA..........................AAAAGCAATGGCTTCCAGC 904
      :                             |||::::::|||    |||
 242 lGluGlnSerThrValGluThrThrThrIleLysAsnGlyGlyGluGlnA 259
```

```
 905 TGGTTCGTAAAGATTGGTTCTATGATGAAATCTTTGCTGGAGATACCCAT 954
      |||   ::::::            |||     :::   |||   |||:::
 259 rgValTyrGluSerArgAlaLeuAspThrThrIleGluGlyGlyThrGln 275

 955 TCAGTATTCTACGAACCACGTCAAAATGGGAAATACTCTTTTAACGAC.. 1002
      |||:::      :::         ::::::  :::   ::::::
 276 SerLeuAsnSerLysSerThrAlaLysAsnThrHisIleTyrSerGlyGl 292

1003 ............GATAATAATGGCACAGGAAAAATCAATGCCAAACATG 1039
                 |||||::::::|||:::    :::         ::::
 292 yThrGlnIleValAspAsnThrSerThrSerAspValIleGluValTyrS 309

1040 AACACAATTCTCTGCCTAATAGA.......................... 1062
      ::   :::   |||     |||
 309 erGlyGlyValLeuAspValArgGlyGlyThrAlaThrAsnValThrGln 325

1063 ..............TTAAAAACACGAACCGTTCAATTGTTTAATGTTTC 1097
                   |||||||||:::|||           ::::::::
 326 HisAspGlyAlaIleLeuLysThrAsnThr.........AsnGlyThrTh 339

1098 TTTATCCGAGACAGCAAGAGAACCTGTTTAT...........CATGCTG 1135
      :::|||   |||      |||    ::::::            |||:::|
 339 rValSerGlyThrAsnSerGluGlyAlaPheSerIleHisAsnHisValA 356

1136 CA...................GGTGGT........GTCAACAGTTAT 1155
      ||                  ||||||        :::|||:::|||
 356 laAspAsnValLeuLeuGluAsnGlyGlyHisLeuAspIleAsnAlaTyr 372

1155 ................................................. 1155

 373 GlySerAlaAsnLysThrIleIleLysAspLysGlyThrMetSerValLe 389

1156 ..................CGACCCAGACTGAATAATGGA.......... 1176
                              |||:::::::||||||
 389 uThrAsnAlaLysAlaAspAlaThrArgIleAspAsnGlyGlyValMetA 406

1177 ..........................................GAAAAT 1182
                                                :::|||
 406 spValAlaGlyAsnAlaThrAsnThrIleIleAsnGlyGlyThrGlnAsn 422

1183 ATTTCCTTTATTGACGAAGGAAAAGGCGAATTGATACTTACCAGCAACAT 1232
      |||:::            :::    |||             :::|||||
 423 IleAsnAsnTyrGlyIleAlaThrGly..............ThrAsnIl 434

1233 CAATCAAGGT...............GCTGGAGGATTATATTTCCAAGGAG 1267
      ||||:::|||               :::||||||          ::::::|
 434 eAsnSerGlyThrGlnAsnIleLysSerGlyGly.........LysAlaA 448

1268 ATTTTACGGTCTCGCCTGAAAATAAC..............GAAACTTGG 1302
      ||   |||:::     :::::::::                :::
 448 spThrThrIleIleSerSerGlySerArgGlnValValGluLysAspGly 464

1303 CAAGGCGCGGGCGTTCATATCAGTGAAGACAGTACCGTTACTTGGAAAGT 1352
        :::   |||   :::||||||      :::::::::           ::
 465 ThrAlaIleGlySerAsnIleSerAlaGlyGlySerLeuIleValTyrTh 481

1353 AAACGGCGTGGCAAACGAC......................CGCCTGT 1378
      ::::|||:::|||:::                           ::::
 481 rGlyGlyIleAlaHisGlyValAsnGlnGluThrGlySerAlaLeuValA 498

1379 CCAAAATCGGCAAAGGCACG......................CTGCAC 1404
      :::::    |||   ||||||                          |||
 498 laAsnThrGlyAlaGlyThrAspIleGluGlyTyrAsnLysLeuSerHis 514

1405 GTTCAAGCCAAAGGG........................GAAAACCAAGG 1430
                  |||                     ||||||   ||
 515 PheThrIleThrGlyGlyGluAlaAsnTyrValValLeuGluAsnThrGl 531

1431 CTCGATCAGCGTGGGCGACGGTACAGTCATTTGGATCAGCAGGCAGACG 1480
      |:::::::::|||           |||        :::           |||
 531 yGluLeuThrValValAlaLysThrSerAlaLysAsnThrThrIleAspT 548

1481 ATAAAGGCAAA............AAACAAGCCTTTAGTGAAATCGGCTTG 1518
         ||||||               |||:::|||   ::::::
 548 hrGlyGlyLysLeuIleValGlnLysGluAlaLysThrAspSerThrArg 564

1519 GTCAGCGGCAGGGGTACGGTGCAACTGAATGCCGAT............. 1554
      :::::::::   |||::::::::::::::::
 565 LeuAsnAsnGlyGlyValLeuGluValGlnAspGlyGlyGluAlaLysHi 581

1555 ....AATCAGTTCAACCCCGACAAACTCTATTTCGGCTTTCGCGGCGGAC 1600
          :::|||   :::          |||     :::       :::|||
 581 sValGluGlnGlnSerGlyGlyAlaLeuIleAlaSerThrThrSerGlyT 598

1601 GTTTG......GATTTAAACGGGCATTCGCTTTCGTTCCACCGTATTCAA 1644
        |||           |||:::::::::   :::|||:::   |||:::
 598 hrLeuIleGluGlyThrAsnSerTyrGlyAspAlaPheTyr...IleArg 613

1645 AATACCGATGAAGGGGCGATGATTGTCAACCACAATCAAGACAAAGAATC 1694
      |||::::::           :::::::::           ::::::    ||
 614 AsnSerGluAlaLysAsnValValLeu.........GluAsnAlaGlySe 627

1695 CACCGTTACCATTACAGGCAATAAAGATATTGCTACAACCGGCAATAACA 1744
      |   :::::::::||||||:::::::   :::   |||      |||    |
 627 rLeuThrValValThrGlySerArgAlaValAspThrIleIleAsnAlaA 644

1745 ACAGCTTGGATAGCAAAAAAGAAATTGCCTACAACGGTTGGTTTGGCGAG 1794
      ||          :::|||   ::::::               :::|||
 644 sn.........GlyLysMetAspVal..............TyrGly... 651

1795 AAAGAT......ACGACCAAAACGAACGGGCGGCTCAACCTTGTTTACCA 1838
      ||||||       |||:::   :::::::::      :::     :::|||
 652 LysAspValGlyThrValLeuAsnSerAlaGlyThrGlnThrIleTyrAl 668

1839 GCCCGCCGCAGAAGACCGCACCCTGCTGCTTTCCGGCGGAACAAATTTAA 1888
         |||::::::|||::::::   :::   :::|||      ::::::
 668 aSerAlaThrSerAspLysAlaAsnIleLysGlyGlyLysGlnThrValT 685

1889 ACGGCAACATCACGCAAACAAAC.........GGCAAACTGTTTTTCAGC 1929
        |||      |||:::::::|||          |||:::   :::   :::
 685 yrGlyLeuAlaThrGluAlaAsnIleGluSerGlyGluGlnIleValAsp 701

1930 GGCAGACCAACACCGCACGCCTACAATCATTTAAACGAC........... 1968
      |||      |||              :::|||:::|||
 702 GlyGlySerThrGluLys......ThrHisIleAsnGlyGlyThrGlnTh 716

1969 ..................................................C 1969
                                                      :
 716 rValGlnAsnTyrGlyLysAlaIleAsnThrAspIleValSerGlyLeuG 733

1970 ATTGGTCGCAAAAAGAGGGCATTCCTCGCGGGGAAATCGTG...TGGGAC 2016
      ::      :::   :::|||      :::|||:::|||:::
 733 lnGlnIleMetAlaAsnGlyThrAlaGluGlySerIleIleAsnGlyGly 749

2017 AACGACTGGATCAACCGCACATTTAAAGCGGAAAACTTCCAAATTAAAGG 2066
      ::::::   :::|||:::            |||||||||      ::::::
 750 SerGlnValValAsnGluGlyGlyLeuAlaGluAsnSerValLeuAsnAs 766

2067 CGGACAGGCGGTGGTTTCCCGCAATGTTGCCAAAGTGAAAGGCGATTGGC 2116
        |||   ::::::      |||:::   :::::::::   |||
 766 pGlyGlyThrLeuAspValArgGluLysGlySerAlaThrGly....... 780

2117 ATTTGAGCAATCACGCCCAAGCAGTTTTTGGTGTCGCACCGCATCAAAGC 2166
         :::::::::   :::|||:::::::::   ::::::   :::   :::
 781 ..IleGlnGlnSerSerGlnGlyAlaLeuValAlaThrThrArgAlaThr 796

2167 CACACAATCTGTACACGTTCGGACTGGACGGGTCTGACAAATTGTGTCGA 2216
      ::::::      ||||||:::|||      |||::::::      :::||
 797 ArgValThrGlyThrArgAlaAsp......GlyValAlaPheSerIleGl 811

2217 AAAA..............................................ACCA 2224
      |:::                                           |||:
 811 uGlnGlyAlaAlaAsnAsnIleLeuLeuAlaAsnGlyGlyValLeuThrV 828

2225 TTACCGACGAT...................................... 2235
```

```
                :: :::|||
 828 alGluSerAspThrSerSerAspLysThrGlnValAsnMetGlyGlyArg 844

2236 .........AAAGTGATTGCTTCATTGACTAAGACCGACATCAGCGGCAA 2276
              |||:::  |||:::  |||  |||  :::::::|||::
 845 GluIleValLysThrLysAlaThrAlaThrGlyThrThrLeuThrGlyGl 861

2277 TGTCGATCTT...........GCCGATCACGCTCATTTAAATCTCACAG 2314
     :  ::::::            |||:::::::::  :::|||      |
 861 yGluGlnIleValGluGlyValAlaAsnGluThrThrIleAsnAspGlyG 878

2315 GGCTTGCCACACTCAACGGCAATCTTAGTGCAAATGGCGAT......... 2355
      ||:::  |||              :::|||||||||||:::
 878 lyIleGlnThr...........ValSerAlaAsnGlyGluAlaIleLys 890

2356 ACACGTTATACAGTCAGCCACAACGCCACCCAAAACGGCAAC........ 2397
      |||:::  :::  :::  :::  |||  |||  |||
 891 ThrLysIleAsnGluGlyGlyThrLeuThrValAsnAspAsnGlyLysAl 907

2398 .CTTAGCCTCGTGGGCAATGCCCAAGCAACATTTAATCAAGCCACATTAA 2446
        ::::::|||  |||:::  |||:::::::::  :::|||   |
 907 aThrAspIleValGlnAsnSerGlyAlaAlaLeuGlnThrSerThrAlaA 924

2447 ACGGCAACACATCGGCTTCGGGCAAT.........GCTTCATTTAATCTA 2487
      |||||     :::  ||||||:::            ::::::|||::::::
 924 snGly...IleGluIleSerGlyThrHisGlnTyrGlyThrPheSerIle 939

2488 AGCGACCACGCCGTACAAAACGGCAGTCTGACGCTTTCCGGCAACGCTAA 2537
      |||   :::  :::  |||    |||    :::||||||
 940 SerGlyAsnLeuAlaThrAsnMetLeuLeuGluAsnGlyGlyAsnLeuLe 956

2538 GGCAAACGTAAGCCATTCCGCACTCAACGGTAATGTCTCCCTAGCCGATA 2587
      :::  ::::::  :::|||  :::::::::|||:::   :::
 956 uValLeuAlaGlyThrGluAlaArgAspSerThrValGlyLysGlyGly. 972

2588 AGGCAGTATTCCATTTTGAAAGCAGCCGCTTTACCGGACAAATCAGCGGC 2637
        |||:::  ::::::  ::::::     |||      ||||||
 973 ..AlaMetGlnAsnLeuGlyGlnAspSerAlaThrLysValAsnSerGly 988

2638 GGCAAGGATACGGCATTACACTTAAAAGACAGCGAATGGACGCTGCCGTC 2687
      |||:::  |||       :::  ||||||:::      :::|||
 989 GlyGlnTyrThrLeuGlyArgSerLysAspGluPheGlnAlaLeuAlaAr 1005

2688 AGGCACGGAATTA..................................... 2700
       :::  :::|||
1005 gAlaGluAspLeuGlnValAlaGlyGlyThrAlaIleValTyrAlaGlyT 1022

2701 ...............................GGCAATTTAAACCTT... 2715
                                    |||:::|||:::|||
1022 hrLeuAlaAspAlaSerValSerGlyAlaThrGlySerLeuSerLeuMet 1038

2716 .........GACAACGCCACC...ATTACACTCAATTCCGCCTATCGC.. 2751
              ||||||:::|||   :::  |||:::::::|||  |||
1039 ThrProArgAspAsnValThrProValLysLeuGluGlyAlaValArgIl 1055

2752 .CACGATGCGGCAGGGGCGCAAACCGGCAGTGCGACAGATGCGCCGCGCC 2800
        |||:::|||            |||:::::::::|||:::
1055 eThrAspSerAlaThrLeuThrLeuGlyAsnGlyValAspThrThrLeuA 1072

2801 GCCGTTCGCGCCGTTCGCGCCGT.......................... 2823
                :::  |||
1072 laAspLeuThrAlaAlaSerArgGlySerValTrpLeuAsnSerAsnAsn 1088

2824 ...................................TCCCTATTA.. 2832
                                          |||||||||
1089 SerCysAlaGlyThrSerAsnCysGluTyrArgValAsnSerLeuLeuLe 1105

2833 ..........................TCCGTTACACCGCCAACTTCGG 2854
                                ::::::::: |||::::::
1105 uAsnAspGlyAspValTyrLeuSerAlaGlnThrAlaAlaProAlaThrT 1122

2855 TAGAATCCCGTTTCAACACGCTGACGGTAAACGGCAAATTGAACGGTCAG 2904
     ::::::::  ::::||||||||||||:::|||   :::|||:::|||:::

1122 hrAsnGlyIleTyrAsnThrLeuThrThrAsn...GluLeuSerGlySer 1137

2905 GGAACATTCCGCTTTATGTCGGAACTCTTCGGCTACCGCAGCGACAAATT 2954
      |||:::|||  :::  :::::::::  |||  |||:::|||:::|||
1138 GlyAsnPheTyrLeuHisThrAsnValAlaGlySerArgGlyAspGlnLe 1154

2955 GAAGCTGGCGGAAAGTTCCGAAGGCACTTACACCTTGGCGGTCAACAATA 3004
      |  :::  :::::::::  |||::::::  :::  |||::::::|
1154 uValValAsnAsnAsnAlaThrGlyAsnPheLysIlePheValGlnAspT 1171

3005 CCGGCAACGAACCTGCAAGCCTCGAACAATTGACGGTAGTGGAAGGAAAA 3054
      |||||  :::|||  |||  :::  :::|||:::|||
1171 hrGlyValSerProGlnSerAspAspAlaMetThrLeuVal......... 1184

3055 GACAACAAACCGCTGTCCGAAAACCTTAATTTCACCCTGCAAAACGAA.. 3102
         |||      :::  :::  :::|||||||||  |||
1185 ......LysThrGlyGlyGlyAspAlaSerPheThrLeuGlyAsnThrGl 1199

3103 ....CACGTCGATGCCGGCGCGTGGCGTTACCAACTCATCCGCAAAGACG 3148
          ||||||  |||:::::::::|||   ::::::::::::|||
1199 yGlyPheValAspLeuGlyThrTyrGluTyr...ValLeuLysSerAspG 1215

3149 GCGAGTTCCGCCTGCATAATCCGGTCAAAGAACAAGAGCTTTCCGACAAA 3198
      || :::|||
1215 ly.AsnSer........................................ 1217

3199 CTCGGCAAGGCAGAAGCCAAAAAACAGGCGGAAAAAGACAACGCGCAAAG 3248
                                                :::
1218 ........................................AsnTrpA 1220

3249 CCTTGACGCGCTGATTGCGGCCGGGCGCGATGCCGTCGAAAAGACAGAAA 3298
        |||||||:::  ::::::|||:::       |||
1220 snLeuThrAsnAspValLysProAsnProAspPro............... 1231

3299 GCGTTGCCGAACCGGCCCGGCAGGCAGGCGGGGAAAATGTCGGCATTATG 3348
       :::||||||
1232 ...IleProAsn...................................... 1234

3349 CAGGCGGAGGAAGAGAAAAAACGGGTGCAGGCGGATAAAGACACCGCCTT 3398

1234 ................................................ 1234

3399 GGCGAAACAGCGCGAAGCGGAAACCCGGCCGGCTACCACCGCCTTCCCCC 3448
                         ||||||      |||   |||:::|||
1235 ................ProLysProAspProLysProAspProLysPro 1245

3449 GCGCCCGCCGCGCCCGCCGGGATTTGCCGCAACTGCAACCCCAACCGCAG 3498
      |||  :::   :::    |||   |||     :::|||   |||
1246 AspPro..AsnProLysProAsp...ProThrProAspProThrProThr 1260

3499 CCCCAACCGCAGCGCGACCTGATCAGCCGTTATGCCAATAGCGGTTTGAG 3548
      |||  |||::::::  :::         ::::::::::::::
1261 ProValProGluLysArgIleThrPro......SerThrAlaAlaValLe 1275

3549 TGAATTTTCCGCCACGCTCAACAGCGTTTTC..............GCCG 3583
      :::::::::|||||||||     ||||||                ::::
1275 uAsnMetAlaAlaThrLeuProLeuValPheAspAlaGluLeuAsnSerI 1292

3584 TACAGGACGAATTAGACCGCGTATTTGCCGAAGACCGCCGCAACGCCGTT 3633
     ::::::::::|||:::  :::  |||:::  ::::::|||  |||
1292 leArgGluArgLeuAsnIleMetLysAlaSerProHisAsnAsnAsnVal 1308

3634 TGGACAAGCGGCATCCGGGACACCAAACACTACCGTTCGCAAGATTTCCG 3683
      |||   ::::::    :::|||:::::    :::   |||
1309 Trp...GlyAlaThrTyrAsnThrArgAsnAsnValThrThrAspAlaGl 1324

3684 CGCCTACCGCCAACAAACCGACCTGCGCCAAATCGGTATGCAGAAAAACC 3733
      |||     :::||||||          ||||||    ::::
1324 yAlaGlyPheGluGlnThr.........LeuThrGlyMetThrValGlyI 1338

3734 TC.......................GGCAGCGGGCGCGTCGGCATCCTG 3759
     ::                        |||          :::|||  :::
1338 leAspSerArgAsnAspIleProGluGlyIleThrThrLeuGlyAlaPhe 1354
```

```
3760 TTTTCGCACAACCGGACCGAAAACACCTTCGACGACGGC........... 3798
      ::::::::::::::::::::          ||||||   |||
1355 MetGlyTyrSerHisSerHisIleGlyPheAspArgGlyGlyHisGlySe 1371

3799 ...............ATCGGCAACTCGGCACGGCTTGCCCACGGCGCCG 3832
                    :::|||:::   |||          |||   :::
1371 rValGlySerTyrSerLeuGlyGlyTyrAlaSerTrpGluHisGluSerG 1388

3833 TTTTC.........GGGCAATACGGCATCGACAGGTTC.........TAC 3864
       |||          |||          ::::::||||||
1388 lyPheTyrLeuAspGlyValValLysLeuAsnArgPheLysSerAsnVal 1404

3865 ATCGGCATCAGCGCGGGCGCGGGTTTAGCAGCGGCAGCCTT...TCAGA 3911
        |||       :::::::::|||   ::::::||||||      |||::
1405 AlaGlyLysMetSerSerGlyGlyAlaAlaAsnGlySerTyrHisSerAs 1421

3912 CGGCATCGGAGGCAAAATCCGCCGCCGCGTGCTGCATTACGGCATTCAGG 3961
      :|||:::||||||                          |||::::
1421 nGlyLeuGlyGlyHis..........................IleGluT 1429

3962 CACGATACCGCGCCGGTTTCGGCGGATTCGGCATCGAACCGCACATCGGC 4011
      ::      |||          |||::::::::::::     |||:::   :::
1429 hrGlyMetArgPheThrAspGlyAsnTrpAsnLeuThrProTyrAlaSer 1445

4012 GCAACGCGCTATTTCGTCCAAAAAGCGGATTACCGCTACGAAAAC..... 4056
        |||    :::   :::::::::    :::|||:::    :::|||
1446 LeuThrGlyPheThrAlaAspAsnProGluTyrHisLeuSerAsnGlyMe 1462

4057 .......GTCAATATCGCCACCCCCGGCCTTGCATTCAACCGCTACCGC. 4098
                 ::::::    |||    ::::::           ||||||
1462 tLysSerLysSerValAspThrArgSerIle............TyrArgG 1475

4099 ..GCGGGCATTAAGGCAGATTATTCATTCAAACCGGCGCAACACATTTCC 4146
          |||          :::|||:::::::::    ::::::    ::::::
1475 luLeuGlyAlaThrLeuSerTyrAsnMetArgLeuGlyAsnGlyMetGlu 1491

4147 ATCACGCCTTATTTGAGCCTGTCCTATACCGATGCCGCTTCGGGCAAAGT 4196
      :::    |||:::|||:::    :::
1492 ValGluProTrpLeuLysAlaAlaValArgLysGluPheValAspAspAs 1508

4197 CCGAACACGCGTCAATACCGCCGTATTGGCTCAGGATTTC......GGCA 4240
        |||::::::||||||:::                 ::::::       |||:
1508 nArgValLysValAsnSerAspGlyAsnPheValAsnTyrLeuSerGlyA 1525

4241 AAACCCGCAGTGCGGAATGGGGCGTAAACGCCGAAATCAAAGGTTTCACG 4290
      ::              :::    |||:::::::|||::::::::::::     |||
1525 rgArgGlyIleTyrGlnAlaGlyIleLysAlaSerPheSerSer...Thr 1540

4291 CTGTCCCTCCACGCTGCCGCCGCCAAAGGCCCGCAACTGGAAGCGCAACA 4340
        ||||||    |||    :::::::::                :::||
1541 LeuSerGlyHisLeuGlyValGly..................TyrSerHi 1551

4341 CAGCGCGGGCATCAAA..................TTAGGCTACCGCTGG 4371
       |||||||||::::::                    |||    :::|||
1551 sSerAlaGlyValGluSerProTrpAsnAlaValAlaGlyValAsnTrp 1567
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01830

```
seq_documentation_block:
ID   AAB01830 standard; Protein; 1222 AA.
XX
AC   AAB01830;
XX
DT   11-SEP-2000  (first entry)
XX
DE   H. influenzae strain K1 mature full-length HMW1A protein, SEQ ID NO:37.
XX
KW   Mature HMW protein; hmw gene; hmwA1; hmwA2; high molecular weight;
KW   non-typeable Haemophilus influenzae; NTHi; non-encapsulated;
KW   recombinant production; Escherichia coli; antibacterial; vaccine;
KW   human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;
KW   detection; diagnosis.
XX
OS   Haemophilus influenzae strain K1.
XX
FH   Key              Location/Qualifiers
FT   Misc-difference  307
FT                    /note= "Encoded by GG"
XX
PN   WO200020609-A2.
XX
PD   13-APR-2000.
XX
PF   07-OCT-1999;   99WO-CA00938.
XX
PR   07-OCT-1998;   98US-0167568.
PR   08-DEC-1998;   98US-0206942.
XX
PA   (CONN-) CONNAUGHT LAB LTD.
XX
PI   Loosmore SM,  Yang Y,  Klein MH;
XX
DR   WPI; 2000-303789/26.
DR   N-PSDB; AAA52180.
XX
PT   Nucleic acid molecule for producing recombinant high molecular weight
PT   proteins of Haemophilus which are used as a vaccine to provide
PT   protection against Haemophilus induced diseases in humans -
XX
PS   Claim 8; Fig 20A-R; 307pp; English.
XX
CC   The invention relates to the recombinant production of Haemophilus
CC   influenzae high molecular weight (HMW) proteins in Escherichia coli. The
CC   expression construct used to effect recombinant expression comprises a
CC   promoter functional in E. coli (e.g., the T7 promoter) operably linked
CC   to a modified hmwABC operon from a non-typeable (non-encapsulated) H.
CC   influenzae (NTHi). Most HMW-expressing NTHi strains contain two hmw gene
CC   clusters termed hmw1ABC and hmw2ABC. Each hmwABC operon comprises hmwA,
CC   hmwB and hmwC genes. The hmwA genes encode the structural HMWA proteins
CC   and the hmwB and hmwC genes encode accessory proteins which are
CC   responsible for post-translational processing and secretion of the HMWA
CC   proteins. The modified hmwABC operon used in the expression construct of
CC   the invention contains an A gene modified such that it encodes only the
CC   mature HMWA. The invention also discloses hmwA genes (AAA52175-A52198)
CC   and HMWA proteins (AAB01824-B01849) from the non-typeable H. influenzae
CC   strains Joyc, K1, K21, LCDC2, PMH1, 15 and 12. The nucleic acids and
CC   vectors are used for the production of recombinant H. influenzae HMW
CC   proteins which can be used as vaccines to mediate a humoral or
CC   cell-mediated immune response to provide protection against diseases in
CC   humans caused by H. influenzae (e.g., otitis media, epiglottitis,
CC   pneumonia and tracheobronchitis). The HMW proteins are also useful as
CC   antigens in immunoassays for detecting antibodies against Haemophilus,
CC   HMW proteins and/or HMW peptides. The nucleotide sequences encoding the
CC   HMW proteins can be used to isolate and clone hmw genes from other
CC   non-typeable strains of Haemophilus via hybridisation reactions. The
CC   present sequence represents a mature HMWA protein from a non-typeable
CC   strain of H. influenzae.
XX
SQ   Sequence   1222 AA;

alignment_scores:
           Quality:  307.00            Length:   1030
             Ratio:  0.594               Gaps:     50
 Percent Similarity:  50.194   Percent Identity:  21.262

alignment_block:
 US-09-303-518D-649 x AAB01830   ..

 Align seg 1/1  to: AAB01830  from: 1  to: 1222

  574 CCTGACCGTGTTCGTATTGGGGCA.........GGCAGGCAATATTGGCG 614
       |||||||:::|||   |||:::|||          ||||||
    1 ProAspAsnValSerIleAsnAlaProAlaLeuGlyArg........... 13
```

```
 615 ATCTGATGAAGATGAGCCCAATAAC......................... 639
             |||:::   ||||||||
  14 ....ThrGluSerThrProAsnAsnAsnGluTyrAspSerProAsnGlnI 29

 639 .................................................. 639

  29 leAsnTyrLysAsnLysProSerLeuSerThrLeuThrAsnThrThrLeu 45

 640 ..............CGCGAAAGTTCATATCATATTGCAAGTGCGTATTC 674
                   |||:::::|||  :::|||:::::::::   ::
  46 GluArgIleLeuLysArgAsnThrSerValAsnIleThrAlaThrLysTh 62

 675 TTGGCTCGTTGGTGGCAAT..........................ACCT 697
      :      |||:::::::::                         |||:
  62 rIleThrValAsnSerAspIleAsnIleGlyAspSerSerHisLeuThrL 79

 698 TTGCACAAAATGGATCAGGTGGTGGCACAGTCAACTTAGGTAGTGAAAAA 747
      ::   :::::|||:::|||   |||   |||||:::
  79 euTrpSerGluGlyGlnGlyArgGlyGlyValAsnVal............ 91

 748 ATTAAACATAGCCCATATGGTTTTTTACCAACAGGAGGCTCATTTGGCGA 797
                                  ||||||:::     :::
  92 ..............................ThrGlyAsnIleThrSerTh 98

 798 CAGTGGCTCACCAATGTTTATCTAT...............GATGCCCAAA 832
      :::::::::   :::   ||||||                |||::::::|
  98 rThrAsnGlyAsnLeuThrIleTyrSerGlyGlyTrpValAspValHisL 115

 833 AGCAAAAGTGGTTAATTAATGGGGTATTGCAAACGGGCAACCCCTATATA 882
      ||                  |||   :::|||::::::|||       |||
 115 ys...............AsnIleThrLeuLysSerGlyTyrLeuAsnIle 126

 883 GGAAAAAGCAATGGCTTCCAGCTGGTTCGTAAAGATTGGTTCTATGATGA 932
            :::::::|||                |||     :::::::::
 127 ThrThrLysGlnGly..................AspIleAlaPheGluAs 137

 933 A.......................ATCTTTGCTGGAGATACCCATTCAG 958
      :                       |||   |||     |||    :::|
 137 pLysProGlyLeuSerAsnLeuThrIleThrAlaLysGlyThrIleAlaV 154

 959 TATTCTACGAACCACGTCAAAATGGGAAATACTCTTTTAACGACGATAAT 1008
      ||                  |||:::|||   :::|||:::   ||||||
 154 al..................AsnAsnLysLysGlyPheArgPheAspAsn 164

1009 .........AATGGCACA...GGAAAAATCAATGCCAAACAT...GAACA 1043
               |||||||||   |||   ::::::   |||:::   |||
 165 ValThrLeuAsnGlyThrGlyGlyGlyLeuSerPheLysTyrIleGluTh 181

1044 CAATTCTCTGCCTAATAGATTAAAAACACGAACCGTTCAATTGTTTAATG 1093
      ::::::        :::::::::::|||:::            :::|||:
 181 rGlyAsnArgAspSerAsnPheGluThrHisPheArgGlyArgLeuAsnI 198

1094 TTTCT..............TTATCCGAGACAGCAAGAGAACCTGTTTAT 1128
      ::|||               |||   ::::::   :::|||
 198 leSerGlyLysValAspIleLeuMetGlnAlaArgGlnGluAsnTrpAsn 214

1129 CATGCTGCAGGTGGTGTCAACAGTTATCGA...CCCAGACTGAATAATGG 1175
      :::        |||   :::   ::::::       |||||||||   ::
 215 ArgArgHisTrpGlyArgSerHisTrpAsnValThrArgLeuAsnValSe 231

1176 AGAAAATATTTCCTTT.........ATTGACGAA............... 1200
      :||||||        |||       ||||||:::
 231 rGluAsnSerTyrPheAsnValThrIleAspSerSerGlySerAlaSerS 248

1201 .....GGAAAAGGCGAATTGATACTTACCAGCAACATCAATCAAGGTGCT 1245
           |||   |||   |||               |||   :::::|||
 248 erProGlyAlaGlyProLeu............AsnAlaGlnSerGlyLeu 260

1246 GGAGGATTATATTTCCAAGGAGAT......................... 1269
      :::|||:::   |||::::::|||
 261 AsnGlyIleSerPheAsnAsnAspThrValPheAsnIleAlaAlaSerSe 277

1270 ..........TTTACGGTCTCGCCT....................GAAA 1288
```

```
             |||::::::::|||                     ::::
 277 rAlaValAsnPheAsnIleLysProProIleValAspLysValThrAsnG 294

1289 ATAACGAAACTTGGCAAGGCGCGGGCGTTCATATCAGTGAAGACAGTACC 1338
      ::|||:::|||               :::::::::   :::      :::
 294 lyAsnHisThrLeuPheLysGlyAsnIleSerValLeuGlyGlyGlyAsp 310

1339 GTTACTTGGAAAGTAAACGGC............................ 1359
      |||:::::      |||:::
 311 ValAsnPheHisPheAsnAlaSerSerSerAsnTyrGlnThrTyrGlyVa 327

1360 ....GTGGCAAACGACCGCCTGTCCAAAATCGGCAAAGGCACGCTGCACG 1405
          :::   ::::::::::|||      |||   ::::::|||
 327 lIleIleGluSerGlnAsnPheSerAlaSerGlyGlySerSerLeuLysP 344

1406 TTCAAGCCAAAGGGGAAAACCAAGGCTCGATCAGCGTGGGCGAC...... 1449
      ::::::::|||::::::::::::::::::::::::::   :::
 344 heLysSerGluGlySerThrHisAlaAlaPheThrIleLysAsnAspLeu 360

1450 .................GGTACAGTCATTTTGGATCAGCAGGCA..... 1476
                       |||::::::   |||:::|||   |||
 361 IleLeuAsnAlaThrGlyGlyAsnIleSerLeuAsnGlnValAlaGlyIl 377

1477 .GACGATAAAGGCAAAAAACAAGCCTTTAGTGAAATCGGCTTGGTCAGCG 1525
       |||::::::   ||||||:::   :::::::::   :::::::::
 377 eAspSerAsnLeuLysLysSerLeuIleAlaAsnLysAsnIleThrPheG 394

1526 GCAGGGGTACGGTGCAACTGAATGCCGATAATCAGTTCAACCCCGACAAA 1575
         |||::::::   |||   |||                     ||||||
 394 luGlyGlyAsnIleThrLeuAlaAla.................AspLys 404

1576 CTCTATTTCGGCTTTCGCGGCGGACGTTTGGATTTAAACGGGCATTCGCT 1625
         :::   :::::|||:::                :::|||   :::::
 405 LysProIleGluIleLysGlyAsnIleThrValLysGluGlyAlaAsnVa 421

1626 TTCGTTCCACCGTATTCAA...AATACCGATGAAGGGGCGATGATTGTCA 1672
      ::::::::::      :::   ::::::|||::::::|||:::   ::::
 421 lThrLeuArgSerAlaAsnTyrGlyAsnAspLysSerAlaLeuSerIleA 438

1673 ACCACAATCAAGACAAAGAATCCACCGTTACCATTACAGGCAAT...... 1716
      ::   |||         :::::::::::::::|||:::|||||||||:::
 438 rgGlyAsnValThrAsnLysGlyAsnLeuThrValThrGlySerAlaIle 454

1717 .........AAAGATATTGCTACAACCGGCAAT................ 1740
            |||::::::::::::   |||:::
 455 AsnIleGluLysAsnLeuThrValGluGlySerAlaLysPheLeuAlaAs 471

1741 ..................................AACAACAGCTTGGATA 1756
                                       :::|||:::     |
 471 nProAsnTyrSerPheAsnValSerGlyLeuPheAspAsnGlnGlyLysS 488

1757 GCAAAAAAGAAATTGCCTACAACGGTTGGTTTGGCGAGAAAGATACGACC 1806
      ||:::     :::||||||   :::|||   :::      ||||||
 488 erAsnIleSerIleAlaLysGlyGlyAlaIle...PheLysAspIleGlu 503

1807 AAAACGAACGGGCGGCTCAACCTTGTTTACCAGCCCGCCGCAGAAGACCG 1856
      :::|||   |||   ||||||:::::   :::         ::::::   ::
 504 AsnThr...GlySerLeuAsnIleThrThrLysSerAspSerAsnHisHi 519

1857 CACCCTGCTGCTTTCCGGCGGAACAAATTTAAACGGCAACATCACGCAAA 1906
      :|||::::::   ::::::   ||||||   :::|||:::::::::   |
 519 sThrIleIleLysGlyAsnIleThrAsnArgLysGlyAspLeuAsnIleT 536

1907 CAAACGGCAAACTGTTTTTCAGCGGCAGACCAACACCGCACGCCTACAAT 1956
      |||||               :::|||      |||
 536 hrAsn...............AsnGlyAspAsnThrGlu.........Ile 544

1957 CATTTAAACGACCATTGGTCGCAAAAAGAGGGCATTCCTCGCGGGGAAAT 2006
      :::::::::   :::      ||||||||||||            |||::::::
 545 GlnIleGlyGlyAsnIleSerGlnLysGlu...........GlyAsnLe 557

2007 CGTGTGGGACAACGACTGGATCAACCGCACATTTAAAGCGGAAAACTTCC 2056
      ::::     ::::::|||   :::|||   |||         |||::::::
```

```
 557 uThrIleSerSerAspLysValAsnIleThr.........GluArgIleT 571

2057 AAATTAAAGGCGGA........................CAGGCGGTGGTT 2082
        |||||||:::|||                          :::    ::::::
 571 hrIleLysAlaGlyValAsnGlyAspAsnSerAspSerAsnGluAlaThr 587

2083 TCCCGCAATGTTGCC..................AAAGTGAAAGGCGATTG 2114
     |||    |||:::::::                  |||:::    :::|||
 588 SerAlaAsnLeuThrIleLysThrLysGluLeuLysLeuThrAsnAspLe 604

2115 GCATTTGAGCAATCACGCCCAAGCAGTTTTGGTGTCGCACCGCATCAAA 2164
      ::::::|||:::       :::|||    :::    :::       :::::::
 604 uAsnIleSerGlyPheAsnLysAlaGluIleThrAlaLysAspAsnSerA 621

2165 GCCACACAATCTGTACACGTTCGGACTGGACGGGTCTGACAAATTGTGTC 2214
      ::    ||||||       :::|||||||    :::|||    |||::::::
 621 snLeuThrIleGlyAspAsnSerAsp...AlaGlyAsnThrAspAlaLys 636

2215 GAAAAAACC.........ATTACCGACGATAAAGTGATTGCTTCA..... 2250
      :::    |||           :::    |||:::|||:::    ||||||
 637 LysValThrPheSerAsnValLysAspSerLysIleSerAlaSerAspHi 653

2251 .............TTGACTAAGACCGACATCAGCGGCAATGTCGATCTTG 2287
                 :::|||::::::    ||||||::::::|||    :
 653 sAsnValThrLeuAsnSerLysValGluThrSerGlyAspThrAspSerT 670

2288 CCGATCACGCTCATTTAAATCTCACAGGGCTT.................. 2319
      :::::    :::        |||    |||||||||
 670 hrGluAspGlyGlyAsnAsnAsnThrGlyLeuThrIleThrAlaLysAsn 686

2320 GCCACACTCAACGGCAATCTTAGTGCAAATGGCGATACACGTTATACAGT 2369
      :::|||:::|||:::|||::::::::::::      ::::::    |||::
 687 ValThrValAsnAsnAsnIleThrSerHisLysThrValAsnIleThrAl 703

2370 CAGCCACAACGCCACCCAAAACGGCAACCTTAGCCTC.........GTGG 2410
      :|||:::|||:::|||    :::::::::    ::::::          :::|
 703 aSerGluAsnValThrThrLysAlaGlyThrThrIleAsnAlaThrThrG 720

2411 GCAATGCCCAAGCAACATTTAATCAAGCCACATTAAACGGC......... 2451
      ||::::::::::::|||    :::    :::    ::::::|||
 720 lySerValGluValThrAlaLysThrGlyAspIleLysGlyGlyIleGlu 736

2452 ..................AACACATCGGCTTCGGGCAATGCTTCATTTAA 2483
                       |||    :::||||||||||:::    ::::::||
 737 SerAsnSerGlyAsnValAsnIleThrAlaSerGlyAsp...ThrLeuAs 752

2484 TCTAAGCGACCACGCCGTACAAAACGGCAGTCTGACGCTTTCCGGCAACG 2533
      |:::|||:::    :::    ||||||    :::::::::    :::::::::|
 752 nValSerAsnIleThrGlyGlnAsnValThrValAlaAlaAlaSerGlyA 769

2534 CTAAGGCAAACGTAAGCCATTCCGCACTCAAC.........GGTAATGTC 2574
      ||    ::::::::::::    |||::::::|||          |||||||:::
 769 laValThrThrThrLysGlySerThrIleAsnAlaThrThrGlyAsnAla 785

2575 TCCCTAGCCGATAAGGCAGTATTCCATTTTGAAAGCAGCCGCTTTACCGG 2624
      :::::::::    |||                              |||||
 786 AsnIleThrThrLys..............................ThrGl 792

2625 ACAAATCAGCGGCGGCAAGGATACGGCA.........TTACACTTAAAAG 2665
      |:::|||:::|||             :::|||          :::::::::
 792 yGluIleAsnGlyGluValLysSerAlaSerGlyAsnValAsnIleThrA 809

2666 ACAGCGAATGGACGCTGCCGTCAGGCACGGAATTAGGCAATTTAAACCTT 2715
      |||         ||||||                  :::::::::|||::::::
 809 laSerGlyAsnThrLeu............AsnValSerAsnIleThrGly 821

2716 GACAACGCCACCATTACACTCAATTCCGCCTATCGCCACGATGCGGCAGG 2765
      :::|||:::|||:::|||    |||||||:::              :::    ||
 822 GlnAsnValThrValThrAlaAsnSerGlyAlaIleThrThrThrGluGl 838

2766 GGCGCAAACCGGCAGTGCGACA...GATGCGCCGCGCCGCCGTTCGCGCC 2812
      |:::          :::::::::|||    ||||||
 838 ySerThrIleAsnAlaThrThrGlyAspAla................... 848

2813 GTTCGCGCCGTTCCCTATTATCCGTTACACCCCCAACTTCGGTAGAATCC 2862

 848 .................................................. 848

2863 CGTTTCAACACGCTGACGGTAAACGGCAAATTGAACGGTCAGGGAACATT 2912
            |||       |||    :::|||::::::||||||:::
 849 ......AsnIleThrThrGlnThrGlyAsnIleAsnGlyLys........ 860

2913 CCGCTTTATGTCGGAACTCTTCGGCTACCGCAGCGACAAATTGAAGCTGG 2962
                                                     :
 861 .................................................V 861

2963 CGGAAAGTTCCGAAGGCACTTACACCTTGGCGGTCAACAAT......... 3003
      ::|||||||||:::|||:::    ||||||    :::::::::
 861 alGluSerSerSerGlySerValThrLeuIleAlaThrGlyGlnThrLeu 877

3004 ...ACCGGCAACGAACCTGCAAGCCTCGAACAATTGACGGTAGTGGAAGG 3050
         :::||||||       ::::::             :::|||:::    :::||
 878 AlaValGlyAsnIleSerGlyAspThrValThrIleThrAlaAspLysGl 894

3051 AAAAGACAACAAACCGCTGTCCGAAAACCTTAATTTCACC 3090
      ||||    :::          |||::::::::::|||    |||
 894 yLysLeuThrThrGlnThrSerSerLysIleAsnGlyThr 907
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01828

seq_documentation_block:

```
ID   AAB01828 standard; Protein; 1228 AA.
XX
AC   AAB01828;
XX
DT   11-SEP-2000  (first entry)
XX
DE   Haemophilus influenzae strain K1 full-length HMW1A protein, SEQ ID NO:34.
XX
KW   HMW protein; hmw gene; hmwA1; hmwA2; high molecular weight;
KW   non-typeable Haemophilus influenzae; NTHi; non-encapsulated;
KW   recombinant production; Escherichia coli; antibacterial; vaccine;
KW   human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;
KW   detection; diagnosis.
XX
OS   Haemophilus influenzae strain K1.
XX
FH   Key                Location/Qualifiers
FT   Misc-difference    313
FT                      /note= "Encoded by GG"
XX
PN   WO200020609-A2.
XX
PD   13-APR-2000.
XX
PF   07-OCT-1999;   99WO-CA00938.
XX
PR   07-OCT-1998;   98US-0167568.
PR   08-DEC-1998;   98US-0206942.
XX
PA   (CONN-) CONNAUGHT LAB LTD.
XX
PI   Loosmore SM,  Yang Y,  Klein MH;
XX
DR   WPI; 2000-303789/26.
DR   N-PSDB; AAA52179.
XX
PT   Nucleic acid molecule for producing recombinant high molecular weight
PT   proteins of Haemophilus which are used as a vaccine to provide
PT   protection against Haemophilus induced diseases in humans -
XX
PS   Claim 12; Fig 20A-R; 307pp; English.
XX
CC   The invention relates to the recombinant production of Haemophilus
CC   influenzae high molecular weight (HMW) proteins in Escherichia coli. The
CC   expression construct used to effect recombinant expression comprises a
```

```
CC   promoter functional in E. coli (e.g., the T7 promoter) operably linked
CC   to a modified hmwABC operon from a non-typeable (non-encapsulated) H.
CC   influenzae (NTHi). Most HMW-expressing NTHi strains contain two hmw gene
CC   clusters termed hmw1ABC and hmw2ABC. Each hmwABC operon comprises hmwA,
CC   hmwB and hmwC genes. The hmwA genes encode the structural HMWA proteins
CC   and the hmwB and hmwC genes encode accessory proteins which are
CC   responsible for post-translational processing and secretion of the HMWA
CC   proteins. The modified hmwABC operon used in the expression construct of
CC   the invention contains an A gene modified such that it encodes only the
CC   mature HMWA. The invention also discloses hmwA genes (AAA52175-A52198)
CC   and HMWA proteins (AAB01824-B01849) from the non-typeable H. influenzae
CC   strains Joyc, K1, K21, LCDC2, PMH1, 15 and 12. The nucleic acids and
CC   vectors are used for the production of recombinant H. influenzae HMW
CC   proteins which can be used as vaccines to mediate a humoral or
CC   cell-mediated immune response to provide protection against diseases in
CC   humans caused by H. influenzae (e.g., otitis media, epiglottitis,
CC   pneumonia and tracheobronchitis). The HMW proteins are also useful as
CC   antigens in immunoassays for detecting antibodies against Haemophilus,
CC   HMW proteins and/or HMW peptides. The nucleotide sequences encoding the
CC   HMW proteins can be used to isolate and clone hmw genes from other
CC   non-typeable strains of Haemophilus via hybridisation reactions. The
CC   present sequence represents an HMWA protein from a non-typeable strain of
CC   H. influenzae.
XX
SQ   Sequence   1228 AA;


alignment_scores:
           Quality:  307.00          Length:   1030
             Ratio:  0.594             Gaps:     50
Percent Similarity:  50.194  Percent Identity: 21.262

alignment_block:
 US-09-303-518D-649 x AAB01828   ..

 Align seg 1/1  to: AAB01828  from: 1  to: 1228

     574 CCTGACCGTGTTCGTATTGGGGCA.........GGCAGGCAATATTGGCG 614
          |||||:::|||  |||:::|||          ||||||
       7 ProAspAsnValSerIleAsnAlaProAlaLeuGlyArg........... 19

     615 ATCTGATGAAGATGAGCCCAATAAC......................... 639
                 |||:::  |||||||||
      20 ....ThrGluSerThrProAsnAsnAsnGluTyrAspSerProAsnGlnI 35

     639 .................................................. 639

      35 leAsnTyrLysAsnLysProSerLeuSerThrLeuThrAsnThrThrLeu 51

     640 ..............CGCGAAAGTTCATATCATATTGCAAGTGCGTATTC 674
                       |||::::::|||  :::|||::::::::   ::
      52 GluArgIleLeuLysArgAsnThrSerValAsnIleThrAlaThrLysTh 68

     675 TTGGCTCGTTGGTGGCAAT..........................ACCT 697
          :     |||:::::::::                          |||:
      68 rIleThrValAsnSerAspIleAsnIleGlyAspSerSerHisLeuThrL 85

     698 TTGCACAAAATGGATCAGGTGGTGGCACAGTCAACTTAGGTAGTGAAAAA 747
          ::   :::::|||:::|||   |||   |||||||:::
      85 euTrpSerGluGlyGlnGlyArgGlyGlyValAsnVal............ 97

     748 ATTAAACATAGCCCATATGGTTTTTTACCAACAGGAGGCTCATTTGGCGA 797
                                          ||||||:::     :::
      98 .............................ThrGlyAsnIleThrSerTh 104

     798 CAGTGGCTCACCAATGTTTATCTAT..............GATGCCCAAA 832
          ::::::::    :::   ||||||              ||::::::|
     104 rThrAsnGlyAsnLeuThrIleTyrSerGlyGlyTrpValAspValHisL 121

     833 AGCAAAAGTGGTTAATTAATGGGGTATTGCAAACGGGCAACCCCTATATA 882
          ||              |||   :::|||:::::|||            |||
     121 ys..............AsnIleThrLeuLysSerGlyTyrLeuAsnIle 132

     883 GGAAAAAGCAATGGCTTCCAGCTGGTTCGTAAAGATTGGTTCTATGATGA 932
            ::::::|||                 |||      ::::::::
     133 ThrThrLysGlnGly.................AspIleAlaPheGluAs 143

     933 A......................ATCTTTGCTGGAGATACCCATTCAG 958
          :                      |||   |||     |||   :::|
     143 pLysProGlyLeuSerAsnLeuThrIleThrAlaLysGlyThrIleAlaV 160

     959 TATTCTACGAACCACGTCAAAATGGGAAATACTCTTTTAACGACGATAAT 1008
          ||                     |||:::|||   :::|||:::   ||||||
     160 al.................AsnAsnLysLysGlyPheArgPheAspAsn 170

    1009 .........AATGGCACA...GGAAAAATCAATGCCAAACAT...GAACA 1043
                   |||||||||   |||   ::::::   |||:::    |||
     171 ValThrLeuAsnGlyThrGlyGlyGlyLeuSerPheLysTyrIleGluTh 187

    1044 CAATTCTCTGCCTAATAGATTAAAAACACGAACCGTTCAATTGTTTAATG 1093
          ::::::      :::::::::::|||:::             :::|||:
     187 rGlyAsnArgAspSerAsnPheGluThrHisPheArgGlyArgLeuAsnI 204

    1094 TTTCT..............TTATCCGAGACAGCAAGAGAACCTGTTTAT 1128
          ::|||             |||   ::::::   :::|||
     204 leSerGlyLysValAspIleLeuMetGlnAlaArgGlnGluAsnTrpAsn 220

    1129 CATGCTGCAGGTGGTGTCAACAGTTATCGA...CCCAGACTGAATAATGG 1175
          :::         |||   :::   ::::::       ||||||||   ::
     221 ArgArgHisTrpGlyArgSerHisTrpAsnValThrArgLeuAsnValSe 237

    1176 AGAAAATATTTCCTTT.........ATTGACGAA................ 1200
          :||||||     |||          |||||||:::
     237 rGluAsnSerTyrPheAsnValThrIleAspSerSerGlySerAlaSerS 254

    1201 .....GGAAAAGGCGAATTGATACTTACCAGCAACATCAATCAAGGTGCT 1245
                |||    |||    |||                |||   ::::::|||
     254 erProGlyAlaGlyProLeu............AsnAlaGlnSerGlyLeu 266

    1246 GGAGGATTATATTTCCAAGGAGAT.......................... 1269
          :::|||:::    |||::::::|||
     267 AsnGlyIleSerPheAsnAsnAspThrValPheAsnIleAlaAlaSerSe 283

    1270 ..........TTTACGGTCTCGCCT....................GAAA 1288
                    |||::::::::|||                     ::::
     283 rAlaValAsnPheAsnIleLysProProIleValAspLysValThrAsnG 300

    1289 ATAACGAAACTTGGCAAGGCGCGGGCGTTCATATCAGTGAAGACAGTACC 1338
          :::|||:::|||              :::::::::   :::       :::
     300 lyAsnHisThrLeuPheLysGlyAsnIleSerValLeuGlyGlyGlyAsp 316

    1339 GTTACTTGGAAAGTAAACGGC............................ 1359
          |||::::::       |||:::
     317 ValAsnPheHisPheAsnAlaSerSerSerAsnTyrGlnThrTyrGlyVa 333

    1360 ....GTGGCAAACGACCGCCTGTCCAAAATCGGCAAAGGCACGCTGCACG 1405
              :::    ::::::::::|||       |||      ::::::|||
     333 lIleIleGluSerGlnAsnPheSerAlaSerGlyGlySerSerLeuLysP 350

    1406 TTCAAGCCAAAGGGGAAAACCAAGGCTCGATCAGCGTGGGCGAC...... 1449
             :::::::::|||::::::::::::::::::::::::    :::
     350 heLysSerGluGlySerThrHisAlaAlaPheThrIleLysAsnAspLeu 366

    1450 ..................GGTACAGTCATTTTGGATCAGCAGGCA..... 1476
                            |||::::::   |||:::|||   |||
     367 IleLeuAsnAlaThrGlyGlyAsnIleSerLeuAsnGlnValAlaGlyIl 383

    1477 .GACGATAAAGGCAAAAAACAAGCCTTTAGTGAAATCGGCTTGGTCAGCG 1525
          |||:::::    ||||||:::   :::::::::    :::::::::
     383 eAspSerAsnLeuLysLysSerLeuIleAlaAsnLysAsnIleThrPheG 400

    1526 GCAGGGGTACGGTGCAACTGAATGCCGATAATCAGTTCAACCCCGACAAA 1575
              |||::::::   |||   |||                  ||||||
     400 luGlyGlyAsnIleThrLeuAlaAla.................AspLys 410

    1576 CTCTATTTCGGCTTTCGCGGCGGACGTTTGGATTTAAACGGGCATTCGCT 1625
```

```
                 :::   ::::::|||:::              :::|||   :::::
 411 LysProIleGluIleLysGlyAsnIleThrValLysGluGlyAlaAsnVa 427

1626 TTCGTTCCACCGTATTCAA...AATACCGATGAAGGGGCGATGATTGTCA 1672
     ::::::::::        :::    :::::||||::::::|||:::    ::::
 427 lThrLeuArgSerAlaAsnTyrGlyAsnAspLysSerAlaLeuSerIleA 444

1673 ACCACAATCAAGACAAAGAATCCACCGTTACCATTACAGGCAAT...... 1716
     ::   |||       :::::::::::::::|||:::||||||||:::
 444 rgGlyAsnValThrAsnLysGlyAsnLeuThrValThrGlySerAlaIle 460

1717 .........AAAGATATTGCTACAACCGGCAAT................ 1740
                  |||::::::::::::    |||:::
 461 AsnIleGluLysAsnLeuThrValGluGlySerAlaLysPheLeuAlaAs 477

1741 ................................AACAACAGCTTGGATA 1756
                                      :::|||:::      |
 477 nProAsnTyrSerPheAsnValSerGlyLeuPheAspAsnGlnGlyLysS 494

1757 GCAAAAAAGAAATTGCCTACAACGGTTGGTTTGGCGAGAAAGATACGACC 1806
     ||:::   :::||||||   :::|||   :::         ||||||
 494 erAsnIleSerIleAlaLysGlyGlyAlaIle...PheLysAspIleGlu 509

1807 AAAACGAACGGGCGGCTCAACCTTGTTTACCAGCCCGCCGCAGAAGACCG 1856
     :::|||   |||   |||||||:::::    :::      ::::::   ::
 510 AsnThr...GlySerLeuAsnIleThrThrLysSerAspSerAsnHisHi 525

1857 CACCCTGCTGCTTTCCGGCGGAACAAATTTAAACGGCAACATCACGCAAA 1906
     :|||::::::   ::::::   ||||||   :::|||:::::::::     |
 525 sThrIleIleLysGlyAsnIleThrAsnArgLysGlyAspLeuAsnIleT 542

1907 CAAACGGCAAACTGTTTTTCAGCGGCAGACCAACACCGCACGCCTACAAT 1956
     |||||              :::|||       |||
 542 hrAsn..............AsnGlyAspAsnThrGlu.........Ile 550

1957 CATTTAAACGACCATTGGTCGCAAAAAGAGGGCATTCCTCGCGGGGAAAT 2006
     :::::::::   :::     |||||||||||             |||:::::
 551 GlnIleGlyGlyAsnIleSerGlnLysGlu...........GlyAsnLe 563

2007 CGTGTGGGACAACGACTGGATCAACCGCACATTTAAAGCGGAAAACTTCC 2056
     ::::   ::::::|||   :::|||    |||         |||::::::
 563 uThrIleSerSerAspLysValAsnIleThr.........GluArgIleT 577

2057 AAATTAAAGGCGGA......................CAGGCGGTGGTT 2082
       ||||||:::|||                     :::    ::::::
 577 hrIleLysAlaGlyValAsnGlyAspAsnSerAspSerAsnGluAlaThr 593

2083 TCCCGCAATGTTGCC.................AAAGTGAAAGGCGATTG 2114
     |||    |||::::::               |||:::     :::|||
 594 SerAlaAsnLeuThrIleLysThrLysGluLeuLysLeuThrAsnAspLe 610

2115 GCATTTGAGCAATCACGCCCAAGCAGTTTTGGTGTCGCACCGCATCAAA 2164
       ::::::|||:::         :::|||     :::   :::        :::::::
 610 uAsnIleSerGlyPheAsnLysAlaGluIleThrAlaLysAspAsnSerA 627

2165 GCCACACAATCTGTACACGTTCGGACTGGACGGGTCTGACAAATTGTGTC 2214
     ::    ||||||      :::||||||    :::|||    |||::::::
 627 snLeuThrIleGlyAspAsnSerAsp...AlaGlyAsnThrAspAlaLys 642

2215 GAAAAAACC.........ATTACCGACGATAAAGTGATTGCTTCA..... 2250
     :::     |||         :::    |||:::|||:::     ||||||
 643 LysValThrPheSerAsnValLysAspSerLysIleSerAlaSerAspHi 659

2251 ............TTGACTAAGACCGACATCAGCGGCAATGTCGATCTTG 2287
                    :::|||:::::    ||||||::::::|||      :
 659 sAsnValThrLeuAsnSerLysValGluThrSerGlyAspThrAspSerT 676

2288 CCGATCACGCTCATTTAAATCTCACAGGGCTT.................. 2319
     :::::    :::         |||     |||||||||
 676 hrGluAspGlyGlyAsnAsnAsnThrGlyLeuThrIleThrAlaLysAsn 692

2320 GCCACACTCAACGGCAATCTTAGTGCAAATGGCGATACACGTTATACAGT 2369
     :::|||:::|||:::|||::::::::::::        ::::::    |||::
 693 ValThrValAsnAsnAsnIleThrSerHisLysThrValAsnIleThrAl 709

2370 CAGCCACAACGCCACCCAAAACGGCAACCTTAGCCTC.........GTGG 2410
     :|||:::|||:::|||    :::::::::    ::::::         :::|
 709 aSerGluAsnValThrThrLysAlaGlyThrThrIleAsnAlaThrThrG 726

2411 GCAATGCCCAAGCAACATTTAATCAAGCCACATTAAACGGC......... 2451
     ||:::::::::::::|||    :::    :::    :::::::|||
 726 lySerValGluValThrAlaLysThrGlyAspIleLysGlyGlyIleGlu 742

2452 ..................AACACATCGGCTTCGGGCAATGCTTCATTTAA 2483
                       |||    :::||||||||||:::      :::::::||
 743 SerAsnSerGlyAsnValAsnIleThrAlaSerGlyAsp...ThrLeuAs 758

2484 TCTAAGCGACCACGCCGTACAAAACGGCAGTCTGACGCTTTCCGGCAACG 2533
     |:::|||:::     :::     ||||||    :::::::::    :::::::::|
 758 nValSerAsnIleThrGlyGlnAsnValThrValAlaAlaAlaSerGlyA 775

2534 CTAAGGCAAACGTAAGCCATTCCGCACTCAAC.........GGTAATGTC 2574
     ||    ::::::::::::     |||::::::|||            |||||||:::
 775 laValThrThrThrLysGlySerThrIleAsnAlaThrThrGlyAsnAla 791

2575 TCCCTAGCCGATAAGGCAGTATTCCATTTGAAAGCAGCCGCTTTACCGG 2624
     :::::::::     |||                           |||||
 792 AsnIleThrThrLys............................ThrGl 798

2625 ACAAATCAGCGGCGGCAAGGATACGGCA.........TTACACTTAAAAG 2665
     |:::|||:::|||            :::|||         :::::::::
 798 yGluIleAsnGlyGluValLysSerAlaSerGlyAsnValAsnIleThrA 815

2666 ACAGCGAATGGACGCTGCCGTCAGGCACGGAATTAGGCAATTTAAACCTT 2715
        |||      ||||||            :::::::::|||::::::
 815 laSerGlyAsnThrLeu............AsnValSerAsnIleThrGly 827

2716 GACAACGCCACCATTACACTCAATTCCGCCTATCGCCACGATGCGGCAGG 2765
     :::|||:::|||:::|||     |||||||:::          :::    ||
 828 GlnAsnValThrValThrAlaAsnSerGlyAlaIleThrThrThrGluGl 844

2766 GGCGCAAACCGGCAGTGCGACA...GATGCGCCGCGCCGCCGTTCGCGCC 2812
     |:::          ::::::::|||    ||||||
 844 ySerThrIleAsnAlaThrThrGlyAspAla.................. 854

2813 GTTCGCGCCGTTCCCTATTATCCGTTACACCGCCAACTTCGGTAGAATCC 2862

 854 .................................................. 854

2863 CGTTTCAACACGCTGACGGTAAACGGCAAATTGAACGGTCAGGGAACATT 2912
            |||      |||    :::|||::::::||||||:::
 855 ......AsnIleThrThrGlnThrGlyAsnIleAsnGlyLys........ 866

2913 CCGCTTTATGTCGGAACTCTTCGGCTACCGCAGCGACAAATTGAAGCTGG 2962
                                                    :
 867 ................................................V 867

2963 CGGAAAGTTCCGAAGGCACTTACACCTTGGCGGTCAACAAT......... 3003
     ::|||||||||:::|||:::    ||||||    :::::::::
 867 alGluSerSerSerGlySerValThrLeuIleAlaThrGlyGlnThrLeu 883

3004 ...ACCGGCAACGAACCTGCAAGCCTCGAACAATTGACGGTAGTGGAAGG 3050
        :::||||||      ::::::           :::|||:::    :::||
 884 AlaValGlyAsnIleSerGlyAspThrValThrIleThrAlaAspLysGl 900

3051 AAAAGACAACAAACCGCTGTCCGAAAACCTTAATTTCACC 3090
     ||||    :::        |||::::::::|||    |||
 900 yLysLeuThrThrGlnThrSerSerLysIleAsnGlyThr 913
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAY81609

```
seq_documentation_block:
ID   AAY81609 standard; Protein; 1237 AA.
XX
AC   AAY81609;
XX
```

```
DT   24-MAY-2000  (first entry)
XX
DE   Streptococcus pneumoniae type 4 protein sequence #109.
XX
KW   Streptococcus pneumoniae; vaccine; screening; protein antigen;
KW   antibacterial; antiinflammatory; meningitis; infection; diagnosis;
KW   pneumococcal disease.
XX
OS   Streptococcus pneumoniae.
XX
PN   WO200006737-A2.
XX
PD   10-FEB-2000.
XX
PF   27-JUL-1999;   99WO-GB02451.
XX
PR   27-JUL-1998;   98GB-0016337.
PR   19-MAR-1999;   99US-0125164.
XX
PA   (MICR-) MICROBIAL TECHNICS LTD.
XX
PI   Gilbert CFG,  Hansbro PM;
XX
DR   WPI; 2000-195300/17.
XX
PT   New Streptococcal protein, useful as a vaccine, for diagnosis of
PT   pneumococcal diseases and for screening agents capable of antagonizing
PT   or inhibiting expression of the protein  -
XX
PS   Claim 1; Page 83; 108pp; English.
XX
CC   AAY81501 to AAY81679 represent specifically claimed protein sequences
CC   isolated from Streptococcus pneumoniae. AAA05407 to AAA05590 represent
CC   specifically claimed nucleotide sequences isolated from S. pneumoniae.
CC   The sequences have antibacterial and antiinflammatory properties.
CC   The protein sequences, and fragments of them, are useful as immunogens
CC   and/or antigens. The nucleotide sequences can be used in vaccines and in
CC   diagnostic assays. The proteins and nucleotides can be useful for the
CC   detection and diagnosis of S. pneumoniae. The protein sequences are also
CC   useful for screening an agent capable of antagonising, inhibiting or
CC   interfering with the function or expression of the proteins in which the
CC   agent is useful for treatment or prophylaxis of S. pneumoniae infection
CC   and meningitis. AAA05591 to AAA05614 represent primers used in the
CC   exemplification of the present invention.
XX
SQ   Sequence   1237 AA;


alignment_scores:
            Quality:   299.00          Length:    1210
              Ratio:   0.514             Gaps:      42
 Percent Similarity:   48.099   Percent Identity:  20.000

alignment_block:
 US-09-303-518D-649 x AAY81609   ..

 Align seg 1/1  to: AAY81609  from: 1  to: 1237

    1273 ACGGTCTCGCCTGAAAATAACGAAACTTGGCAAGGCGCGGGCGTTCATAT 1322
          |||   :::   |||:::   :::|||       :::|||:::::   ||
     103 ThrSerAlaSerGluSerAlaSerThr......SerAlaSerThrSerIl 117

    1323 CAGTGAAGACAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAAACGACC 1372
          ||||   :::|||||||||:::   :::   :::::::::|||::::::
     117 eSerAlaSerSerThrValValGlySerGlnThrAlaAlaAlaThrGluA 134

    1373 GCCTGTCCAAAATCGGCAAAGGCACGCTGCACGTTCAAGCCAAAGGGGAA 1422
                :::|||                  |||:::         :::
     134 laThrAlaLys..................LysValGluGluAspArgLys 144

    1423 AACCAAGGCTCGATCAGCGTGGGCGACGGTACAGTCATTTGGATCAGCA 1472
          :::    :::|||       |||::::::   |||   :::   :::   ||
     145 LysProAlaSerAspTyrValAlaSerValThrAsnVal...AsnLeuGl 160

    1473 GGCAGACGATAAAGGCAAAAAACAAGCC...................... 1500
          |:::        |||     :::|||::::::
     160 nSerTyrAlaLysArgArgLysArgSerValAspSerIleGluGlnLeuL 177

    1501 .........................TTTAGTGAAATCGGCTTGGTCAGC 1524
                                  ||||||          :::|||:::
     177 euAlaSerIleLysAsnAlaAlaValPheSerGlyAsnThrIleValAsn 193

    1525 GGCAGGGGTACGGTGCAACTGAATGCCGATAATCAGTTCAACCCCGACAA 1574
          |||      :::::::::   :::   :::            :::      ||
     194 GlyAlaProAlaIleAsnAlaSerLeuAsnIleAlaLysSerGluThrLy 210

    1575 ACTCTATTTCGGCTTTCGCGGCGGACGTTTGGATTTAAACGGGCATTCGC 1624
          |:::|||              |||         :::|||          :::
     210 sValTyr.........ThrGlyGluGlyValAspSerValTyrArgValP 224

    1625 TTTCGTTCCACCGTATTCAAAATACCGATGAAGGGGCGATGATTGTCAAC 1674
               :::::::::::::::   |||::::::|||:::   ::::::
     224 roIleTyrTyrLysLeuLysValThrAsnAspGlySerLysLeuThr... 239

    1675 CACAATCAAGACAAAGAATCCACCGTTACCATTACA...GGCAATAAAGA 1721
                                    |||   |||:::|||      |||
     240 ..................PheThrTyrThrValThrTyrValAsnProLy 250

    1722 TATTGCTACAACCGGCAATAACAACAGCTTGGATAGCAAAAAAGAAATTG 1771
                         ||||||   :::|||:::               :::|||
     250 sThrAsnAspLeuGlyAsnIleSerSerMetArgProGlyTyrSerIleT 267

    1772 CCTACAACGGTTGG.................................... 1785
             :::|||
     267 yrAsnSerGlyThrSerThrGlnThrMetLeuThrLeuGlySerAspLeu 283

    1786 .........TTTGGCGAGAAAGAT...ACGACCAAAACGAACGGGCGGCT 1823
                  |||   |||:::         |||      ||||||||
     284 GlyLysProSerGlyValLysAsnTyrIleThrAspLysAsnGlyArgGl 300

    1824 CAACCTTGTTTACCAGCCCGCCGCAGAAGACCGCACCCTGCTGCTTTCCG 1873
             |||   |||:::   :::            |||:::        :::|
     300 nValLeuSerTyrAsnThrSer............ThrMetThrThrGlnG 313

    1874 GCGGAACAAATTTAAACGGCAACATCACGCAAACAAACGGCAAACTGTTT 1923
          ||:::            ||||||   :::|||   ||||||      |||
     313 lySerGlyTyrThrTrpGlyAsnGlyAlaGlnMetAsnGly......Phe 327

    1924 TTCAGCGGCAGACCA................................... 1938
          |||:::   :::
     328 PheAlaLysLysGlyTyrGlyLeuThrSerSerTrpThrValProIleTh 344

    1939 .........................ACACCGCACGCC............. 1950
                                  ||||||:::|||
     344 rGlyThrAspThrSerPheThrPheThrProTyrAlaAlaArgThrAspA 361

    1951 ........TACAATCATTTAAACGACCATTGGTCGCAAAAAGAGGGCATT 1992
                  |||::::::|||
     361 rgIleGlyIleAsnTyrPheAsnGly........................ 369

    1993 CCTCGCGGGGAAATCGTGTGGGACAACGACTGGATCAACCGCACATTTAA 2042
             |||::::::|||   ::::::            :::         ::
     370 ...GlyGlyLysValValGluSerSerThrThrSerGlnSerLeuSerGl 385

    2043 AGCGGAAAACTTCCAAATTAAAGGCGGACAGGCGGTGGTTTCCCGCAATG 2092
          ::::::::::::::::::::::::::::|||::::::   :::
     385 nSerLysSerLeuSerValSerAlaSerGlnSerAlaSerAlaSerAlaS 402

    2093 TTGCCAAAGTGAAAGGCGATTGGCATTTGAGCAATCACGCC......... 2133
           ::::::::::::::::::         |||      |||
     402 erThrSerAlaSerAlaSerAlaSerThrSerAlaSerAlaSerAlaSer 418

    2134 ..........................CAAGCAGTTTTGGTGTCGCACC 2156
                                   :::|||      :::::::::
     419 ThrSerAlaSerAlaSerAlaSerThrSerAlaSerValSerAlaSerTh 435
```

```
2157 GCATCAAAGCCACACAATCTGTACACGTTCGGACTGGACGGGTCTGACAA 2206
                 |||    :::         |||    ::::::    ::::::   |||:
 435 rSerAlaSerAlaSerAlaSerThrSerAlaSerAlaSerAlaSerThrS 452

2207 ATTGTGTCGAAAAAACCATTACCGACGATAAAGTGATTGCTTCATTG... 2253
     :::::    |||::::::   |||:::    ::::::   ||||||
 452 erAlaSerGluSerAlaSerThrSerAlaSerAlaSerAlaSerThrSer 468

2254 ..................ACTAAGACCGACATCAGCGGCAATGTCGATCT 2285
                       |||:::::::::    |||::::::::::::
 469 AlaSerAlaSerAlaSerThrSerAlaSerAlaSerAlaSerThrSerAl 485

2286 TGCCGATCACGCTCATTTAAATCTCACAGGGCTTGCCACACTCAACGGCA 2335
       ::::::     |||        :::    ::::::   |||:::  :::::::
 485 aSerGluSerAlaSerThrSerAlaSerAlaSerAlaSerThrSerAlaS 502

2336 ATCTTAGTGCAAATGGCGATACACGTTATACAGTCAGCCACAACGCCACC 2385
     ::     ||||||:::    ::::::      :::::::|||    :::|||:::
 502 erGluSerAlaSerThrSerAlaSerAlaSerAlaSerThrSerAlaSer 518

2386 CAAAACGGCAACCTTAGCCTCGTGGGCAATGCCCAAGCAACATTTAAT.. 2433
       :::::::::    |||        |||:::|||:::::::::     :::
 519 AlaSerAlaSerThrSerAlaSerGlySerAlaSerThrSerThrSerAl 535

2434 .............CAAGCCACATTAAACGGCAACACATCGGCTTCGGGCA 2470
                  :::|||:::    :::::::::||||||||||||::::
 535 aSerAlaSerThrSerAlaSerAlaSerAlaSerThrSerAlaSerAlaS 552

2471 ATGCTTCATTTAATCTAAGCGAC............CACGCCGTACAAAAC 2508
     :::||||||:::::    |||:::                |||    ::::::
 552 erAlaSerIleSerAlaSerGluSerAlaSerThrSerAlaSerGluSer 568

2509 GGCAGTCTGACGCTTTCCGGCAACGCTAAGGCAAACGTAAGCCATTCCGC 2558
     :::|||    :::    |||::::::|||::::::::::::|||:::|||||
 569 AlaSerThrSerThrSerAlaSerAlaSerThrSerAlaSerGluSerAl 585

2559 ACTCAACGGTAATGTCTCCCTAGCCGATAAGGCAGTATTCCATTTTGAAA 2608
     |    ::::::       :::    |||:::    ::::::                  |
 585 aSerThrSerAlaSerAlaSerAlaSerThrSerAlaSerAlaSerAlaS 602

2609 GCAGCCGCTTTACCGGACAAATCAGCGGCGGCAAGGATACGGCATTACAC 2658
     ||:::    :::::::::    |||    :::    :::::::::
 602 erThrSerAlaSerAlaSerAlaSerThrSerAlaSerAlaSerThrSer 618

2659 TTAAAAGACAGCGAATGGACGCTGCCGTCAGGCACGGAATTAGGCAATTT 2708
       ::::::|||      |||      |||::::::           :::
 619 AlaSerGluSerAlaSerThrSerAlaSerAlaSerAlaSerThrSerAl 635

2709 AAACCTTGACAACGCCACCATTACACTCAATTCCGCCTATCGCCACGATG 2758
       :::    :::    :::|||    :::       ||||||          :
 635 aSerAlaSerAlaSerThrSerAlaSerAlaSerAlaSerThrSerAlaS 652

2759 CGGCAGGGGCGCAAACCGGCAGTGCGACAGATGCGCCGCGCCGCGTTCG 2808
      ::|||:::|||:::|||:::::::::::::::|||            |||
 652 erAlaSerAlaSerThrSerAlaSerValSerAlaSerThrSerAlaSer 668

2809 CGCCGTTCGCGCCGTTCCCTATTATCCGTTACACCGCCAACTTCGGTAGA 2858
          :::       |||       :::    :::       ::::::   ||
 669 AlaSerAlaSerThrSerAlaSerAlaSerAlaSerThrSerAlaSerGl 685

2859 ATCCCGTTTCAACACGCTGACGGTAAACGGCAAATTGAACGGTCAGGGAA 2908
     ||||        ::::::    :::::::::::::::    :::::::::::::
 685 uSerAlaSerThrSerAlaSerAlaSerAlaSerThrSerAlaSerAlaS 702

2909 CATTCCGCTTTATGTCGGAACTCTTCGGCTACCGCAGCGACAAATTGAAG 2958
     ::                ::::::      :::     |||    :::    :::
 702 er............AlaSerThrSerAlaSerAlaSerAlaSerThrSer 714

2959 CTGGCGGAAAGTTCCGAAGGCACTTACACCTTGGCGGTCAACAATACCGG 3008
         :::||||||:::::    :::    :::    ::::::::::::::::
 715 AlaSerGluSerAlaSerThrSerAlaSerAlaSerAlaSerThrSerAl 731

3009 CAACGAACCTGCAAGCCTCGAACAATTGACGGTAGTGGAAGGAAAAGACA 3058
     ::::|||    ||||||    :::       :::    ::::::    :::    :
 731 aSerGluSerAlaSerThrSerAlaSerAlaSerAlaSerThrSerAlaS 748

3059 ACAAACCGCTGTCCGAAAACCTTAATTTCACCCTGCAAAACGAACACGTC 3108
     ::            |||    :::    ::: :::    ::::::|||:::|||||
 748 erAlaSerAlaSerThrSerAlaSer.GlySerAlaSerThrSerThrSe 764

3109 GATGCCGGCGCGTGGCGTTACCAACTCATCCGCAAAGACGGCGAGTTCCG 3158
     |         |||             ::::::::||||||:::|||::::::||||
 764 rAlaSerAlaSerThrSerAlaSerAlaSerAlaSerThrSerAlaSerA 781

3159 CCTGCATAATCCGGTCAAAGAACAAGAGCTTTCCGACAAACTCGGCAAG. 3207
     ||                |||    ::::::|||        |||::::::::::
 781 laSerAlaSerIleSerAlaSerGluSerAlaSerThrSerAlaSerGlu 797

3208 ....................GCAGAAGCCAAAAAACAGGCGGAAAAAGA 3236
                         |||:::|||:::    :::|||:::::::
 798 SerAlaSerThrSerThrSerAlaSerAlaSerThrSerAlaSerGluSe 814

3237 CAACGCGCAAAGCCTTGACGCGCTGATTGCGGCCGGGCGCGATGCCGTCG 3286
     :    :::    |||    :::|||       :::::::::    :::|||
 814 rAlaSerThrSerAlaSerAlaSerAlaSerAlaSerAlaSerAlaSerA 831

3287 AAAAGACAGAAAGCGTTGCC............................. 3306
       :::|||::::::    |||
 831 laSerThrSerAlaSerAlaSerAlaArgGlnValArgArgProGlnPro 847

3307 ..................GAACCGGCCCGGCAGGCAGGCGGGGAAAATGT 3338
                       :::|||:::||||||            ::::::||
 848 ValHisLeuAsnArgHisGlnProValArgGlnPro......GlnGlnVa 862

3339 CGGCATTATGCAGGCGGAGGAAGAGAAAAAAACGG............... 3372
     |    :::    |||    ::::::::::::    :::
 862 lLeuValHisGlnLeuHisGlnArgValHisArgLeuGlnHisGlnP 879

3373 ...........................GTGCAGGCGGATAAAGACACC 3393
                                 |||    ::::::
 879 roValProArgLeuGlnArgGlnProValArgGlnLeuGlnGlnValPro 895

3394 GCCTTGGCGAAACAGCGCGAAGCGGAAACCCGGCCGGCTACCACCGCCTT 3443
      :::|||    :::|||::::::          :::|||
 896 ValLeuGlnSerGlnHisGlnGlnValLeuGlnProGlnHisArgGlnVa 912

3444 CCCCCGCGCCCGCCGCGCC.................CGCCGGGATTTGC 3475
      ||||||    ::::::|||                   ||||||:::     |
 912 lProArgLeuGlnGlnAlaHisGlnHisLeuAsnGlnArgArgGlnAlaP 929

3476 CGCAACTGCAACCCCAACCGCAGCCCCAACCGCAGCGCGA.......... 3515
     |||||||||||    |||         ||||||||||||
 929 roGlnLeuGlnGlnValProValArgGlnProGlnArgArgGlnValArg 945

3516 ...CCTGATCAGCCGTTATGCCAATAGCGGTTTGAGTGAATTTTCCGCCA 3562
        |||:::|||    |||    :::
 946 GlnProGlnGlnValLeuValHis......................... 953

3563 CGCTCAACAGCGTTTTCGCCGTACAGGACGAATTAGACCGCGTATTTGCC 3612
         |||    :::    :::|||:::    |||:::|||
 954 ....GlnLeuGlnHisGlnArgValHisArgLeuArg............ 964

3613 GAAGACCGCCGCAACGCCGTTTGGACAAGCGGCATCCGGGACACCAAACA 3662
                                 |||:::|||    ||||||:::
 965 ...........................ArgGlnProValHisGlnSer 971

3663 .....................CTACCGTTCGCAAGATTTCCGCGC..... 3686
                           ||||||    :::    ||||||
 972 GlnGlnValProValArgGlnLeuProHisGlnGlnValProArgLeuGl 988

3687 ...............................CTACCGCCAACAAACCGAC 3705
                                    |||    |||    :::|
 988 nGlnAlaProValArgArgLeuGlnGlnValLeuAlaProGlnProGlnP 1005

3706 CTGCGCCAAATCGGTATGCAGAAAAACCTCGGCAGCGGGCGCGTCGGCAT 3755
      ||    |||    |||          ::::::        ||||||    :::||||||
```

```
1005 roGlnProValArgGlnProGlnGlnValSerGlnArgLeuAsnArgHis 1021

3756 CCT.......................GTTTTCGCACAACCGGACCGAAA 3781
                               :::     ||||||:::||||::
1022 GlnArgValArgProLeuGlnGlnValLeuAlaProGlnProGlnArgGl 1038

3782 ACACCTTCGACGACGGCATCGGCAACTCGGCACGGCTTGCCCACGGCGCC 3831
     ::::::::|||   :::||||||                    ::::::|
1038 nGlnValHisArgLeuGlnArgGlnArgValArgLeuAsnArgHisGlnA 1055

3832 GTTTTCGGGCAATACGG....................CATCGACAGGTT 3860
     ||   |||   ::::::                    :::||||||
1055 rgValArgProLeuGlnGlnValLeuAlaProGlnProGlnArgGlnGln 1071

3861 CTACATCGG......CATCAGCGCGGGCGCGGGTTTTAGCAG........ 3896
     :::||||||       ||||||||   |||    :::   |||
1072 ValHisArgLeuGlnHisGlnArgValArgProLeuGlnGlnValLeuAl 1088

3897 .CGGCAGCCTTTCAGACGG......CATCGGAGGCAAAATCCGCCGCCGC 3939
        ||||||   |||:::     ||||||   |||           |
1088 aProGlnProGlnArgGlnGlnValHisArgLeuGln............A 1101

3940 GTGCTGCATTACGGCATTCAGGCACGATACCGCGCCGGTTTCGGCGGATT 3989
     ||       :::|||   |||                 :::   :::|||:::
1101 rgGlnArgValArgLeuSer...............GlnHisGlnArgVal 1112

3990 CGGCAT...CGAACCGCACATCGGCGCAACGCGCTATTTCGTCCAAAAAG 4036
     |||:::   :::   ||||||:::         |||   :::
1113 ArgGlnProGlnGlnAlaHisGlnLeuLeuAsnLeuHisGln........ 1126

4037 CGGATTACCGCTACGAAAACGTCAATATCGCCACCCCCGGCCTTGCATTC 4086
            |||:::|||:::   |||:::|||:::      |||    :::|
1127 .......ProValArgGlnProGlnHisArgGln...AlaProGlnLeuG 1140

4087 AACCGCTACCGCGCGGGCATTAAGGCAGATTATTCATTCAAACCGGCGCA 4136
     ||    :::|||
1140 lnGlnValPro....................................... 1143

4137 ACACATTTCCATCACGCCTTATTTGAGCCTGTCCTATACCGATGCCGCTT 4186

1143 .................................................. 1143

4187 CGGGCAAAGTCCGAACACGCGTCAATACCGCCGTATTGGCTCA...GGAT 4233
        ::::::|||:::    ||||||    |||||||:::   :::
1144 .ValArgGlnProGlnArgArgGlnValArgArgLeuGlnGlnValProV 1160

4234 TTCGGCAAAACCCGCAGTGCGGAATGGGGCGTAAACGCCGAAATCAAAGG 4283
        ||||||      |||       :::   :::    :::|||:::|||
1160 alArgGlnProGlnGlnValProValArgGlnProGlnArgArgGlnVal 1176

4284 TTTCACGCTGTCCCTCCACGCTGCCGCCGCCAAAGGCCCGCAACTGGAAG 4333
        :::        |||    :::     :::|||:::::|||:::
1177 ArgArgProGlnProValHisLeuAsnArgHisGlnProValArgGlnPr 1193

4334 CGCAACACAGCGCGGGCATCAAATTAGG 4361
                       ||||||:::::
1193 oGlnGlnValLeuValHisGlnLeuGln 1202
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:AAR26503

```
seq_documentation_block:
ID   AAR26503 standard; Protein; 911 AA.
XX
AC   AAR26503;
XX
DT   12-MAR-1993  (first entry)
XX
DE   prn proteins.
XX
KW   B. bronchiseptica; P.68; outer membrane protein; piglet; probe;
KW   atrophic rhinitis; alternative cleavage.
XX
OS   Bordetalla bronchiseptica.
XX
FH   Key             Location/Qualifiers
FT   Protein         35..632
FT                   /label= P.68
FT   Region          266..279
FT                   /label= Repeat_region
FT   Region          570..589
FT                   /label= Repeat_region
FT   Peptide         260..262
FT                   /label= RGD_tripeptide
FT   Peptide         701..703
FT                   /label= RGD_tripeptide
XX
PN   WO9217587-A.
XX
PD   15-OCT-1992.
XX
PF   27-MAR-1992;   92WO-GB00561.
XX
PR   27-MAR-1991;   91GB-0006568.
XX
PA   (WELL ) WELLCOME FOUND LTD.
XX
PI   Charles IG;
XX
DR   WPI; 1992-366258/44.
DR   N-PSDB; AAQ34566.
XX
PT   DNA encoding a Bordetella bronchiseptica protein - used for
PT   obtaining vaccines for preventing respiratory diseases, partic.
PT   atrophic rhinitis in pigs
XX
PS   Claim 1; Fig 1; 28pp; English.
XX
CC   The sequence given is the P.94 antigen from B. bronchiseptica.  The
CC   P.68 antigen is formed by alternative cleavage of this protein.
CC   P.68 is an outer membrane protein with a molecular weight of 68 kD
CC   which is associated with protection of piglets against atrophic
CC   rhinitis.  The DNA sequence encoding these proteins was derived by
CC   standard recombinant DNA techniques using P.68 probes to isolate the
CC   entire P.94 sequence.
XX
SQ   Sequence   911 AA;

alignment_scores:
           Quality:  297.50          Length:   1079
             Ratio:  0.613             Gaps:     50
Percent Similarity:  44.949   Percent Identity: 21.965

alignment_block:
 US-09-303-518D-649 x AAR26503   ..

 Align seg 1/1  to: AAR26503  from: 1  to: 911

  1417 GGGGAAAACCAAGGCTCGATCAGCGTGGGCGACGGTACAGTCATTTGGA 1466
        ||||||:::|||    :::|||   :::
    45 GlyGluArgGlnHisGlyIleHisIle....................... 53

  1467 TCAGCAGGCAGACGATAAAGGCAAAAAACAAGCCTTTAGTGAAATCGGCT 1516
        :::|||:::|||        |||   :::   |||   :::
    54 .LysGlnSerAspGlyAlaGlyValArgThrAlaThrGlyThrThrIleL 70

  1517 TGGTCAGCGGCAGGGGTACGGTGCAACTGAATGCCGATAATCAGTTCAAC 1566
          ||||||||||||     :::        :::       :::|||
    70 ysValSerGlyArgGlnAlaGlnGlyValLeuLeuGluAsn......... 83

  1567 CCCGACAAACTCTATTTCGGCTTTCGCGGCGGACGTTTGGATTTAAACGG 1616
                 :::   |||::::::|||   :::       :::||
    84 ...ProAlaAlaGluLeuArgPheGlnAsnGlySerValThrSerSerGl 99

  1617 GCATTCGCTTTCGTTCCACCGTATTCAAAATACCGATGAAGGGGCGATGA 1666
         |                       |||      |||||||||:::
```

```
  99 y.......................GlnLeuPheAspGluGlyVal.... 106

1667 TTGTCAACCACAATCAAGACAAAGAATCCACCGTTACCATTACAGGCAAT 1716
                    :::       :::|||||||||:::    ::::::
 107 ..............ArgArgPheLeuGlyThrValThrValLysAlaGly 118

1717 AAA.........GATATTGCTACAACCGGCAATAACAACAGCTTGGATAG 1757
     |||         |||  ||||||   :::|||
 119 LysLeuValAlaAspHisAlaThrLeuAlaAsn................. 129

1758 CAAAAAAGAAATTGCCTACAACGGTTGGTTTGGCGAGAAAGATACGACCA 1807
                                    :::||||||
 130 ..................................ValSerAspThrArgA 135

1808 AAACGAACGGGCGGCTCAACCTTGTT............TACCAGCCCGCC 1845
         :::|||          |||             |||   :::
 135 spAspAspGlyIleAlaLeuTyrValAlaGlyGluGlnAlaGlnAlaSer 151

1846 GCAGAAGACCGCACCCTGCTGCTTTCCGGCGGAACAAATTTA......AA 1889
        |||    ||||||        :::||||||:::::::::       ::
 152 IleAlaAspSerThrLeuGlnGlyAlaGlyGlyValArgValGluArgGl 168

1890 CGGCAACATCACGCAAACAAACGGCAAACTGTTTTTCAGCGGC....... 1932
     ::::|||:::|||       ::::::   :::     :::|||
 168 yAlaAsnValThrValGlnArgSerThrIleAlaAspGlyGlyLeuHisI 185

1933 ...........AGACCAACACCGCACGCCTACAATCATTTAAACGACCAT 1971
               :::|||
 185 leGlyThrLeuGlnProLeu.............................. 191

1972 TGGTCGCAAAAAGAGGGCATTCCTCGCGGGGAAATCGTGTGGGACAACGA 2021
           |||   |||   :::|||   :::::::::|||      :::
 192 ......GlnProGluAspLeuProProSerArgValValLeuGlyAspTh 206

2022 CTGGATCAAC................................CGCACAT 2038
       ::::::                                     :::|
 206 rSerValThrAlaValProAlaSerGlyAlaProAlaAlaValSerValP 223

2039 TTAAAGCGGAAAACTTCCAAATTAAAGGCGGACAGGCGGTGGTTTCCCGC 2088
      ||  |||:::::::::   :::    ||||||:::   :::    :::|||
 223 heGlyAlaAsnGluLeuThrValAspGlyGlyHisIleThrGlyGlyArg 239

2089 AATGTTGCCAAAGTGAAAGGCGATTGGCATTTGAGCAATCACGCCCAAGC 2138
        ::::::                                  :::   ||
 240 AlaAlaGly..............................ValAlaAl 245

2139 AGTTTTTGGTGTCGCACCGCATCAAAGCCACACAATCTGTACACGTTCGG 2188
     |:::   |||:::       |||    :::::::::        |||:::|
 245 aMetAspGlyAlaIleValHisLeuGlnArgAlaThrIleArgArgGlyA 262

2189 AC................................................ 2190
      ||
 262 spAlaProAlaGlyGlyAlaValProGlyGlyAlaValProGlyGlyPhe 278

2191 ..................TGGACGGGTCTGACAAATTGTGTCGAAAAAAC 2222
                       |||    |||:::
 279 GlyProLeuLeuAspGlyTrpTyrGlyVal................... 288

2223 CATTACCGACGATAAAGTGATTGCTTCATTGACTAAGACCGACATCAGC. 2271
                                                |||:::|||
 289 ......................................AspValSerA 292

2272 ..GGCAATGTCGATCTTGCCGATCACGCTCATTTAAATCTCACAGGGCTT 2319
       ::::::||||||||||||:::              :::::::::
 292 spSerThrValAspLeuAlaGln...........SerIleValGluAla 304

2320 GCCACACTCAACGGCAATCTTAGTGCAAATGGCGATACACGTTATACAGT 2369
              |||::::::   :::   |||:::        :::|||   |||||
 305 ProGlnLeuGlyAlaAlaIleArgAlaGlyArgGlyAlaArgValThrVa 321

2370 CAGCCACAAC......GCCACCCAAAACGGCAACCTTAGCCTCGTGGGCA 2413
      ||||    :::       ::::::   :::||||||:::     :::|||:
 321 lSerGlyGlySerLeuSerAlaProHisGlyAsnValIleGluThrGlyG 338

2414 ATGCCCAAGCAACATTTAAT..................CAAGCCACATTA 2445
     :::::          |||                     :::   ||||||
 338 lyGlyAlaArgArgPheProProProAlaSerProLeuSerIleThrLeu 354

2446 AACGGCAACACATCGGCTTCGGGCAATGCT...TCATTTAATCTAAGCGA 2492
     ::::::::::::   |||:::|||:::|||      :::::::::
 355 GlnAlaGlyAlaArgAlaGlnGlyArgAlaLeuLeuTyrArgValLeuPr 371

2493 CCACGCCGTACAAAACGGCAGTCTGACGCTTTCCGGCAACGCTAAGGCAA 2542
      :::    |||:::          |||||||||:::|||:::|||::::::
 371 oGluProValLys.........LeuThrLeuAlaGlyGlyAlaGlnGly. 384

2543 ACGTAAGCCATTCCGCACTCAACGGTAATGTCTCCCTAGCCGATAAGGCA 2592
                         :::|||:::::         :::
 385 ....................GlnGlyAspIleValAlaThr......... 391

2593 GTATTCCATTTTGAAAGCAGCCGCTTTACCGGACAAATCAGCGGCGGCAA 2642
                 |||         :::   |||      ||||||
 392 ............GluLeuProProIleProGlyAlaSerSerGlyProLe 404

2643 GGATACGGCATTACACTTAAAAGACAGCGAATGGACGCTGCCGTCAGGCA 2692
      |||:::||||||       ::::::::::::||||||      :::
 404 uAspValAlaLeu...AlaSerGlnAlaArgTrpThr...GlyAlaThrA 419

2693 CGGAATTAGGCAATTTAAACCTTGACAACGCCACCATTACACTCAATTCC 2742
         :::    :::|||::::::||||||||||||    ::::::
 419 rgAlaValAspSerLeuSerIleAspAsnAlaThrTrpValMet...... 433

2743 GCCTATCGCCACGATGCGGCAGGGGCGCAAACCGGCAGTGCGACAGATGC 2792
                                                 ||||||
 434 ..........................................ThrAsp.. 435

2793 GCCGCGCCGCCGTTCGCGCCGTTCGCGCCGTTCCCTATTATCCGTTACAC 2842
                         :::|||:::   :::   |||   ::::::
 436 ...................AsnSerAsnValGlyAlaLeuArgLeuAlaS 446

2843 CGCCAACTTCGGTAGAA....................TCCCGTTTCAAC 2871
           |||||||:::                     :::||||||:::
 446 erAspGlySerValAspPheGlnGlnProAlaGluAlaGlyArgPheLys 462

2872 ACGCTGACGGTAAACGGCAAATTGAACGGTCAGGGAACATTCCGCTTTAT 2921
       |||   |||:::       |||   |||:::|||   ||||||:::
 463 CysLeuMetValAsp...ThrLeuAlaGlySerGlyLeuPheArgMetAs 478

2922 GTCGGAACTCTTCGGCTACCGCAGCGACAAATTGAAGCTGGCGGAAAGTT 2971
                       ||||||||||||   :::    :::::::
 478 nValPheAlaAspLeuGlyLeuSerAspLysLeuValValMetArgAspA 495

2972 CCGAAGGCACTTACACCTTGGCGGTCAACAATACCGGCAACGAACCTGCA 3021
     ::::::|||   :::   |||   |||:::|||:::|||:::||||||||
 495 laSerGlyGlnHisArgLeuLeuValArgAsnSerGlySerGluProAla 511

3022 AGCCTCGAACAATTGACGGTAGTGGAAGGAAAAGACAACAAACCGCTGTC 3071
      |||   :::   :::   :::|||:::                |||   ::
 512 SerGlyAsnThrMetLeuLeuValGln...........ThrProArgGl 524

3072 CGAAAACCTTAATTTCACCCTGCAAAACGAA......CACGTCGATGCCG 3115
      ::::     :::|||||||||   |||:::         ||||||    |
 524 ySerAlaAlaThrPheThrLeuAlaAsnLysAspGlyLysValAspIleG 541

3116 GCGCGTGGCGTTACCAACTCATCCGCAAAGACGGCGAGTTCCGCCTGCAT 3165
      ||::::::||||||:::|||
 541 lyThrTyrArgTyrArgLeu............................. 547

3166 AATCCGGTCAAAGAACAAGAGCTTTCCGACAAACTCGGCAAGGCAGAAGC 3215

 547 ................................................. 547

3216 CAAAAAACAGGCGGAAAAAGACAACGCGCAAAGCCTTGACGCGCTGATTG 3265
                |||   :::   |||:::|||          :::|||:::
 548 ..........AlaAlaAsnGlyAsnGlyGlnTrp......SerLeuValG 559
```

```
3266 CGGCCGGGCGCGATGCCGTCGAAAAGACAGAAAGCGTTGCCGAACCGGCC 3315
      ::|||   |||  |||:::                        :::||
 559 lyAla.LysAlaProProAlaPro....................LysPr 568

3316 CGGCAGGCAGGCGGGGAAAATGTCGGCATTATGCAGGCGGAGGAAGAGAA 3365
      |:::   |||  |||
 568 oAlaProGlnProGlyPro............................. 574

3366 AAAACGGGTGCAGGCGGATAAAGACACCGCCTTGGCGAAACAGCGCGAAG 3415

 574 ................................................. 574

3416 CGGAAACCCGGCCGGCTACCACCGCCTTCCCCCGCGCCCGCCGCGCCCGC 3465
        :::||||||     ||||||   ||||||:::|||||||:::||||
 575 ...GlnProGlyProGlnProProGln.ProProGlnProProGlnProP 590

3466 CGGGATTTGCCGCAACTGCAACCCCAACCGCAGCCCCAACCGCAGCGCGA 3515
      ||           ||| |||   |||||||||         |||::
 590 roGlnArgGlnProGluAla.ProAlaProGlnProProAlaGlyArgGl 606

3516 CCTGATCAGCCGTTATGCCAATAGCGGTTTGAGTGAATTTTCCGCCACGC 3565
      :|||   |||     ||||||::::::::::::      ::::::    |
 606 uLeu...SerAlaAlaAlaAsnAlaAlaValAsnThrGlyGlyValGlyL 622

3566 TCAACAGCGTT......TTCGCCGTACAGGACGAATTAGACCGCGTATTT 3609
      ||  |||:::     :::|||   ::::::  |||::::::  :::
 622 euAlaSerThrLeuTrpTyrAlaGluSerAsnAlaLeuSerLysArgLeu 638

3610 GCCGAAGACCGCCGCAAC...........GCCGTTTGGACAAGCGGCAT 3647
      :::|||   |||   |||            ::::::|||     |||::
 639 GlyGluLeuArgLeuAsnProAspAlaGlyGlyAlaTrpGlyArgGlyPh 655

3648 CCGGGACACCAAACACTACCGTTCGCAAGATTTCCGCGCCTACCGCCAAC 3697
      :   :::  ::::::        ::::::      |||   :::   |||:
 655 eAlaGlnArgGlnGlnLeuAspAsnArgAlaGlyArgArgPheAspGlnL 672

3698 AAACCGACCTGCGCCAAATCGGTATGCAGAAAAACCTC............ 3735
      ::::          ::::::|||  :::     :::
 672 ysValAlaGlyPheGluLeuGlyAlaAspHisAlaValAlaValAlaGly 688

3736 GGCAGCGGGCGCGTCGGCATCCTGTTTTCGCACAACCGGACCGAAAACAC 3785
      |||      ::::::|||   |||   :::::::::|||   ::::::
 689 GlyArgTrpHisLeuGlyGlyLeuAlaGlyTyrThrArgGlyAspArgGl 705

3786 CTTC...GACGACGGCATCGGCAACTCGGCACGGCTTGCCCACGGCGCCG 3832
       |||         ||||||   |||::::::  ::::::    |||:::
 705 yPheThrGlyAspGlyGlyGlyHisThrAspAsnValAsnValGlyGlyT 722

3833 TTTTCGGGCAATACGGCATCGACAGGTTCTACATCGGCATCAGCGCGGGC 3882
                      :::   :::   ||||||:::      :::
 722 yrProThrTyrIleAlaAsnSerGlyPheTyrLeuAspAlaThrLeuArg 738

3883 GCGGGTTTTAGCAGCGGCAGCCTT...........TCAGACGGC..... 3915
      |||:::      ::::::::::::              |||||||||
 739 AlaSerArgLeuGluAsnAspPheLysValAlaGlySerAspGlyTyrAl 755

3916 .ATCGGAGGCAAAATCCGCCGCCGCGTGCTGCATTACGGCATTCAGGCAC 3964
       :::   ||||||  |||                 :::||||:::   |||
 755 aValLysGlyLysTyrArgThr............HisGlyValGlyAlaS 768

3965 GATACCGCGCCGGT...............TTCGGCGGATTCGGCATCGAA 3999
        :::||||||                    |||:::   :::|||
 768 erLeuGluAlaGlyArgArgPheAlaHisAlaAspGlyTrpPheLeuGlu 784

4000 CCGCACATCGGCGCAACGCGCTATTTCGTCCAAAAAGCGGATTACCGCTA 4049
      |||:::        :::   :::   |||   :::::::::||||||
 785 ProGlnAlaGluLeuAlaValPheArgValGlyArgGlySerTyrArgAl 801

4050 CGAAAAC...GTCAATATCGCCACCCCCGGCCTTGCATTCAACCGCTACC 4096
         |||   :::::::::           |||   :::         |
 801 aAlaAsnGlyLeuArgValArgAspGluGlyGlySerSerValLeuGlyA 818

4097 GCGCGGGCATTAAGGCAGATTATTCATTCAAACCGGCGCAACACATTTCC 4146
      ||   |||:::::::::          ::::::   |||       :::
 818 rgLeuGlyLeuGluValGlyLysArgIleGluLeuAlaGlyGlyArgGln 834

4147 ATCACGCCTTATTTGAGCCTGTCCTATACCGATGCCGCTTCGGGCAAAGT 4196
      :::   ||||||:::::   |||      :::      :::|||
 835 ValGlnProTyrIleLysAlaSerValLeuGlnGluPheAspGlyAlaGl 851

4197 CCGAACACGCGTCAATACCGCCGTATTGGCTCAGGATTTCGGCAAAACCC 4246
         :::|||:::|||      :::        ::::::       ||||
 851 yThrValArgThrAsnGlyIleAlaHisArgThrGluLeuArgGlyThrA 868

4247 GCAGTGCGGAATGGGGCGTAAACGCCGAAATCAAAGGTTTCACGCTGTCC 4296
      ||    ||||||   |||                          |||:::
 868 rg...AlaGluLeuGly...........................LeuGly 874

4297 CTCCACGCTGCCGCCGCCAAAGGCCCGCAACTGGAAGCG......CAACA 4340
      :::    ||||||   :::::::|||   :::|||   |||       :::::
 875 MetAlaAlaAlaLeuGlyArgGlyHisSerLeuTyrAlaSerTyrGluTy 891

4341 CAGCGCGGGCATCAAATTAGGCTACCGCTGG 4371
      :|||   |||   |||||||:::      |||
 891 rSerLysGlyProLysLeuAlaMetProTrp 901
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:AAR14320

```
seq_documentation_block:
ID   AAR14320 standard; Protein; 911 AA.
XX
AC   AAR14320;
XX
DT   20-JAN-1992  (first entry)
XX
DE   Pertactin antigen P.68.
XX
KW   Pertactin; Pichia; B. pertussis; B. parapertussis.
XX
OS   Bordetella bronchiseptica.
XX
FH   Key             Location/Qualifiers
FT   Peptide         266..270
FT                   /label= repeat
FT   Peptide         271..275
FT                   /label= repeat
FT   Peptide         570..572
FT                   /label= repeat
FT   Peptide         574..576
FT                   /label= repeat
FT   Peptide         578..580
FT                   /label= repeat
FT   Peptide         581..583
FT                   /label= repeat
FT   Peptide         584..586
FT                   /label= repeat
FT   Peptide         587..589
FT                   /label= repeat
FT   Peptide         599..601
FT                   /label= repeat
XX
PN   WO9115571-A.
XX
PD   17-OCT-1991.
XX
PF   28-MAR-1991;   91WO-GB00487.
XX
PR   02-APR-1990;   90GB-0007416.
XX
PA   (WELL ) WELLCOME FOUNDATION LTD.
XX
PI   Clare JJ,  Romanos MA;
XX
DR   WPI; 1991-325214/44.
DR   N-PSDB; AAQ14319.
XX
```

```
PT   Pichia microorganism transformants - for production of
PT   Bordetella pertactin antigens for whooping cough vaccines
XX
PS   Disclosure; Fig 1B; 38pp; English.
XX
CC   Pichia microorganisms are transformed for the expression of
CC   pertactin antigens.  DNA sequence used are represented in AAQ14319-20
CC   encoding the B. bronchiseptica P.68 and B. parapertussis P.70 antigen
CC   respectively or the B. pertussis P.69 encoding sequence described
CC   by I.G. Charles et al. Proc. Natl. Acad. Sci. USA, Vol. 80:3554-3448
CC   (1989).
XX
SQ   Sequence   911 AA;


alignment_scores:
          Quality:  295.50          Length:   1079
            Ratio:  0.612             Gaps:     50
 Percent Similarity:  44.764   Percent Identity:  21.965

alignment_block:
 US-09-303-518D-649 x AAR14320   ..

 Align seg 1/1  to: AAR14320  from: 1  to: 911

   1417 GGGGAAAACCAAGGCTCGATCAGCGTGGGCGACGGTACAGTCATTTTGGA 1466
         ||||||:::|||  :::|||  :::
     45 GlyGluArgGlnHisGlyIleHisIle....................... 53

   1467 TCAGCAGGCAGACGATAAAGGCAAAAAACAAGCCTTTAGTGAAATCGGCT 1516
          :::|||:::|||     |||   :::   |||   :::
     54 .LysGlnSerAspGlyAlaGlyValArgThrAlaThrGlyThrThrIleL 70

   1517 TGGTCAGCGGCAGGGGTACGGTGCAACTGAATGCCGATAATCAGTTCAAC 1566
           ||||||||||||   :::      :::      :::|||
     70 ysValSerGlyArgGlnAlaGlnGlyValLeuLeuGluAsn......... 83

   1567 CCCGACAAACTCTATTTCGGCTTTCGCGGCGGACGTTTGGATTTAAACGG 1616
                 :::   |||:::::|||   :::     :::||
     84 ...ProAlaAlaGluLeuArgPheGlnAsnGlySerValThrSerSerGl 99

   1617 GCATTCGCTTTCGTTCCACCGTATTCAAAATACCGATGAAGGGGCGATGA 1666
        |                        |||       |||||||||:::
     99 y.......................GlnLeuPheAspGluGlyVal.... 106

   1667 TTGTCAACCACAATCAAGACAAAGAATCCACCGTTACCATTACAGGCAAT 1716
                       :::    ::|||||||||:::   ::::::
    107 .............ArgArgPheLeuGlyThrValThrValLysAlaGly 118

   1717 AAA.........GATATTGCTACAACCGGCAATAACAACAGCTTGGATAG 1757
        |||         |||   ||||||   :::|||
    119 LysLeuValAlaAspHisAlaThrLeuAlaAsn................ 129

   1758 CAAAAAAGAAATTGCCTACAACGGTTGGTTTGGCGAGAAAGATACGACCA 1807
                                         :::||||||
    130 ...............................ValSerAspThrArgA 135

   1808 AAACGAACGGGCGGCTCAACCTTGTT...........TACCAGCCCGCC 1845
             :::|||         |||              |||   :::
    135 spAspAspGlyIleAlaLeuTyrValAlaGlyGluGlnAlaGlnAlaSer 151

   1846 GCAGAAGACCGCACCCTGCTGCTTTCCGGCGGAACAAATTTA......AA 1889
                 |||   ||||||   :::|||||::::::::       ::
    152 IleAlaAspSerThrLeuGlnGlyAlaGlyGlyValArgValGluArgGl 168

   1890 CGGCAACATCACGCAAACAAACGGCAAACTGTTTTTCAGCGGC....... 1932
          ::::|||:::|||      ::::::   :::     :::|||
    168 yAlaAsnValThrValGlnArgSerThrIleValAspGlyGlyLeuHisI 185

   1933 ...........AGACCAACACCGCACGCCTACAATCATTTAAACGACCAT 1971
                   :::|||
    185 leGlyThrLeuGlnProLeu............................. 191

   1972 TGGTCGCAAAAAGAGGGCATTCCTCGCGGGGAAATCGTGTGGGACAACGA 2021
              |||  |||   :::|||   :::::::::|||     :::
    192 ......GlnProGluAspLeuProProSerArgValValLeuGlyAspTh 206

   2022 CTGGATCAAC..............................CGCACAT 2038
          ::::::                                   :::|
    206 rSerValThrAlaValProAlaSerGlyAlaProAlaAlaValSerValP 223

   2039 TTAAAGCGGAAAACTTCCAAATTAAAGGCGGACAGGCGGTGGTTTCCCGC 2088
        ||   |||:::::::   :::    ||||||:::   :::    :::|||
    223 heGlyAlaAsnGluLeuThrValAspGlyGlyHisIleThrGlyGlyArg 239

   2089 AATGTTGCCAAAGTGAAAGGCGATTGGCATTTGAGCAATCACGCCCAAGC 2138
           ::::::                                 :::   ||
    240 AlaAlaGly..............................ValAlaAl 245

   2139 AGTTTTGGTGTCGCACCGCATCAAAGCCACACAATCTGTACACGTTCGG 2188
        |:::   |||:::       |||   :::::::::          |||:::|
    245 aMetAspGlyAlaIleValHisLeuGlnArgAlaThrIleArgArgGlyA 262

   2189 AC................................................ 2190
        ||
    262 spAlaProAlaGlyGlyAlaValProGlyGlyAlaValProGlyGlyPhe 278

   2191 ..................TGGACGGGTCTGACAAATTGTGTCGAAAAAAC 2222
                          |||   |||:::
    279 GlyProLeuLeuAspGlyTrpTyrGlyVal.................... 288

   2223 CATTACCGACGATAAAGTGATTGCTTCATTGACTAAGACCGACATCAGC. 2271
                                               |||:::|||
    289 .....................................AspValSerA 292

   2272 ..GGCAATGTCGATCTTGCCGATCACGCTCATTTAAATCTCACAGGGCTT 2319
          ::::::||||||||||||:::            :::::::::
    292 spSerThrValAspLeuAlaGln...........SerIleValGluAla 304

   2320 GCCACACTCAACGGCAATCTTAGTGCAAATGGCGATACACGTTATACAGT 2369
             |||::::::   :::   |||:::      :::|||   |||||
    305 ProGlnLeuGlyAlaAlaIleArgAlaGlyArgGlyAlaArgValThrVa 321

   2370 CAGCCACAAC......GCCACCCAAAACGGCAACCTTAGCCTCGTGGGCA 2413
        ||||   :::      ::::::   :::|||||||:::     :::|||:
    321 lSerGlyGlySerLeuSerAlaProHisGlyAsnValIleGluThrGlyG 338

   2414 ATGCCCAAGCAACATTTAAT..................CAAGCCACATTA 2445
         :::::         |||                    :::   ||||||
    338 lyGlyAlaArgArgPheProProProAlaSerProLeuSerIleThrLeu 354

   2446 AACGGCAACACATCGGCTTCGGGCAATGCT...TCATTTAATCTAAGCGA 2492
         ::::::::::::   |||:::|||:::|||      :::::::::
    355 GlnAlaGlyAlaArgAlaGlnGlyArgAlaLeuLeuTyrArgValLeuPr 371

   2493 CCACGCCGTACAAAACGGCAGTCTGACGCTTTCCGGCAACGCTAAGGCAA 2542
          :::   |||:::               |||||||||:::|||:::|||::::::
    371 oGluProValLys.........LeuThrLeuAlaGlyGlyAlaGlnGly. 384

   2543 ACGTAAGCCATTCCGCACTCAACGGTAATGTCTCCCTAGCCGATAAGGCA 2592
                              :::|||::::::    :::
    385 ...................GlnGlyAspIleValAlaThr......... 391

   2593 GTATTCCATTTGAAAGCAGCCGCTTTACCGGACAAATCAGCGGCGGCAA 2642
                      |||       :::   |||     ||||||
    392 ...........GluLeuProProIleProGlyAlaSerSerGlyProLe 404

   2643 GGATACGGCATTACACTTAAAAGACAGCGAATGGACGCTGCCGTCAGGCA 2692
          |||:::||||||        ::::::::::::||||||      :::
    404 uAspValAlaLeu...AlaSerGlnAlaArgTrpThr...GlyAlaThrA 419

   2693 CGGAATTAGGCAATTTAAACCTTGACAACGCCACCATTACACTCAATTCC 2742
             :::   :::|||::::::||||||||||||    ::::::
    419 rgAlaValAspSerLeuSerIleAspAsnAlaThrTrpValMet...... 433

   2743 GCCTATCGCCACGATGCGGCAGGGGCGCAAACCGGCAGTGCGACAGATGC 2792
```

```
                                                   ||||||
 434 .........................................ThrAsp.. 435

2793 GCCGCGCCGCCGTTCGCGCCGTTCGCGCCGTTCCCTATTATCCGTTACAC 2842
                         :::|||:::  :::  |||  ::::::
 436 ...................AsnSerAsnValGlyAlaLeuArgLeuAlaS 446

2843 CGCCAACTTCGGTAGAA.....................TCCCGTTTCAAC 2871
         |||||||:::                          :::|||||||:::
 446 erAspGlySerValAspPheGlnGlnProAlaGluAlaGlyArgPheLys 462

2872 ACGCTGACGGTAAACGGCAAATTGAACGGTCAGGGAACATTCCGCTTTAT 2921
        |||   |||:::    |||  |||:::|||  |||||||:::
 463 CysLeuMetValAsp...ThrLeuAlaGlySerGlyLeuPheArgMetAs 478

2922 GTCGGAACTCTTCGGCTACCGCAGCGACAAATTGAAGCTGGCGGAAAGTT 2971
                          |||||||||||   :::   :::::::
 478 nValPheAlaAspLeuGlyLeuSerAspLysLeuValValMetArgAspA 495

2972 CCGAAGGCACTTACACCTTGGCGGTCAACAATACCGGCAACGAACCTGCA 3021
     ::::::|||   :::   |||  |||:::|||:::|||:::|||||||||
 495 laSerGlyGlnHisArgLeuLeuValArgAsnSerGlySerGluProAla 511

3022 AGCCTCGAACAATTGACGGTAGTGGAAGGAAAAGACAACAAACCGCTGTC 3071
     |||   :::   :::   :::|||:::                |||   ::
 512 SerGlyAsnThrMetLeuLeuValGln...........ThrProArgGl 524

3072 CGAAAACCTTAATTTCACCCTGCAAAACGAA......CACGTCGATGCCG 3115
     ::::         :::|||||||||  |||:::          ||||||   |
 524 ySerAlaAlaThrPheThrLeuAlaAsnLysAspGlyLysValAspIleG 541

3116 GCGCGTGGCGTTACCAACTCATCCGCAAAGACGGCGAGTTCCGCCTGCAT 3165
     ||::::::||||||:::|||
 541 lyThrTyrArgTyrArgLeu.............................. 547

3166 AATCCGGTCAAAGAACAAGAGCTTTCCGACAAACTCGGCAAGGCAGAAGC 3215

 547 .................................................. 547

3216 CAAAAAACAGGCGGAAAAAGACAACGCGCAAAGCCTTGACGCGCTGATTG 3265
                |||   :::   |||:::|||          :::|||:::
 548 ..........AlaAlaAsnGlyAsnGlyGlnTrp......SerLeuValG 559

3266 CGGCCGGGCGCGATGCCGTCGAAAAGACAGAAAGCGTTGCCGAACCGGCC 3315
     ::|||    |||   |||:::                        :::||
 559 lyAla.LysAlaProProAlaPro....................LysPr 568

3316 CGGCAGGCAGGCGGGGAAAATGTCGGCATTATGCAGGCGGAGGAAGAGAA 3365
     |:::   |||   |||
 568 oAlaProGlnProGlyPro.............................. 574

3366 AAAACGGGTGCAGGCGGATAAAGACACCGCCTTGGCGAAACAGCGCGAAG 3415

 574 .................................................. 574

3416 CGGAAACCCGGCCGGCTACCACCGCCTTCCCCCGCGCCCGCCGCGCCCGC 3465
         :::||||||      ||||||   ||||||:::||||||:::||||
 575 ...GlnProGlyProGlnProProGln.ProProGlnProProGlnProP 590

3466 CGGGATTTGCCGCAACTGCAACCCCAACCGCAGCCCCAACCGCAGCGCGA 3515
     ||             |||  |||   |||||||||        |||::
 590 roGlnArgGlnProGluAla.ProAlaProGlnProProAlaGlyArgGl 606

3516 CCTGATCAGCCGTTATGCCAATAGCGGTTTGAGTGAATTTTCCGCCACGC 3565
     :|||   |||        ||||||::::::::::::       ::::::   |
 606 uLeu...SerAlaAlaAlaAsnAlaAlaValAsnThrGlyGlyValGlyL 622

3566 TCAACAGCGTT......TTCGCCGTACAGGACGAATTAGACCGCGTATTT 3609
     ||   |||:::         :::|||   ::::::   |||::::::   :::
 622 euAlaSerThrLeuTrpTyrAlaGluSerAsnAlaLeuSerLysArgLeu 638

3610 GCCGAAGACCGCCGCAAC...........GCCGTTTGGACAAGCGGCAT 3647
     :::|||   |||   |||                :::::::|||      |||::

 639 GlyGluLeuArgLeuAsnProAspAlaGlyGlyAlaTrpGlyArgGlyPh 655

3648 CCGGGACACCAAACACTACCGTTCGCAAGATTTCCGCGCCTACCGCCAAC 3697
     :   :::    ::::::         ::::::      |||   :::   |||:
 655 eAlaGlnArgGlnGlnLeuAspAsnArgAlaGlyArgArgPheAspGlnL 672

3698 AAACCGACCTGCGCCAAATCGGTATGCAGAAAAACCTC............ 3735
     :::::        ::::::|||    :::       :::
 672 ysValAlaGlyPheGluLeuGlyAlaAspHisAlaValAlaValAlaGly 688

3736 GGCAGCGGGCGCGTCGGCATCCTGTTTTCGCACAACCGGACCGAAAACAC 3785
     |||       ::::::|||   |||   :::::::::|||   ::::::
 689 GlyArgTrpHisLeuGlyGlyLeuAlaGlyTyrThrArgGlyAspArgGl 705

3786 CTTC...GACGACGGCATCGGCAACTCGGCACGGCTTGCCCACGGCGCCG 3832
      |||        ||||||   |||::::::         :::      |||:::
 705 yPheThrGlyAspGlyGlyGlyHisThrAspSerValHisValGlyGlyT 722

3833 TTTTCGGCAATACGGCATCGACAGGTTCTACATCGGCATCAGCGCGGGC 3882
                 :::   :::   ||||||:::        :::
 722 yrAlaThrTyrIleAlaAsnSerGlyPheTyrLeuAspAlaThrLeuArg 738

3883 GCGGGTTTTAGCAGCGGCAGCCTT...........TCAGACGGC..... 3915
     |||:::      ::::::::::::           |||||||||
 739 AlaSerArgLeuGluAsnAspPheLysValAlaGlySerAspGlyTyrAl 755

3916 .ATCGGAGGCAAAATCCGCCGCCGCGTGCTGCATTACGGCATTCAGGCAC 3964
      :::   ||||||   |||                 :::|||:::   |||
 755 aValLysGlyLysTyrArgThr...........HisGlyValGlyAlaS 768

3965 GATACCGCGCCGGT..............TTCGGCGGATTCGGCATCGAA 3999
        :::||||||                         |||:::   :::|||
 768 erLeuGluAlaGlyArgArgPheAlaHisAlaAspGlyTrpPheLeuGlu 784

4000 CCGCACATCGGCGCAACGCGCTATTTCGTCCAAAAAGCGGATTACCGCTA 4049
     |||:::         :::   :::   |||      :::::::||||||
 785 ProGlnAlaGluLeuAlaValPheArgValGlyGlyGlySerTyrArgAl 801

4050 CGAAAAC...GTCAATATCGCCACCCCCGGCCTTGCATTCAACCGCTACC 4096
          |||   :::::::::               |||   :::          |
 801 aAlaAsnGlyLeuArgValArgAspGluGlyGlySerSerValLeuGlyA 818

4097 GCGCGGGCATTAAGGCAGATTATTCATTCAAACCGGCGCAACACATTTCC 4146
     ||   |||:::::::::            ::::::   |||         :::
 818 rgLeuGlyLeuGluValGlyLysArgIleGluLeuAlaGlyGlyArgGln 834

4147 ATCACGCCTTATTTGAGCCTGTCCTATACCGATGCCGCTTCGGGCAAAGT 4196
     :::   ||||||::::::   |||       :::         :::|||
 835 ValGlnProTyrIleLysAlaSerValLeuGlnGluPheAspGlyAlaGl 851

4197 CCGAACACGCGTCAATACCGCCGTATTGGCTCAGGATTTCGGCAAAACCC 4246
         :::|||:::|||      :::          ::::::      ||||
 851 yThrValArgThrAsnGlyIleAlaHisArgThrGluLeuArgGlyThrA 868

4247 GCAGTGCGGAATGGGGCGTAAACGCCGAAATCAAAGGTTTCACGCTGTCC 4296
     ||    ||||||   |||                             |||:::
 868 rg...AlaGluLeuGly.........................LeuGly 874

4297 CTCCACGCTGCCGCCGCCAAAGGCCCGCAACTGGAAGCG......CAACA 4340
     :::      ||||||   ::::::|||   :::|||   |||       :::::
 875 MetAlaAlaAlaLeuGlyArgGlyHisSerLeuTyrAlaSerTyrGluTy 891

4341 CAGCGCGGGCATCAAATTAGGCTACCGCTGG 4371
     :|||   |||   |||||||:::        |||
 891 rSerLysGlyProLysLeuAlaMetProTrp 901
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:AAR25578

seq_documentation_block:
ID   AAR25578 standard; Protein; 922 AA.
XX
AC   AAR25578;
XX

```
DT   08-JAN-1993  (first entry)
XX
DE   Bordetella parapertussis P95 antigen precursor.
XX
KW   Whooping cough; P70 antigen; P95 precursor protein; vaccination.
XX
OS   Bordetella parapertussis.
XX
FH   Key             Location/Qualifiers
FT   Protein         35..643
FT                   /label= P70
FT   Binding-site    260..262
FT                   /note= "motif associated with cell-cell adhesion"
FT   Region          266..285
FT                   /note= "contains 5 direct, tandem repeats"
FT   Region          575..612
FT                   /note= "contains 9 direct repeats of Pro-Gln-Pro"
FT   Binding-site    712..714
FT                   /note= "motif associated with cell-cell adhesion"
XX
PN   WO9211292-A.
XX
PD   09-JUL-1992.
XX
PF   23-DEC-1991;   91WO-GB02302.
XX
PR   21-DEC-1990;   90GB-0027901.
XX
PA   (WELL ) WELLCOME FOUND LTD.
XX
PI   Charles IG;
XX
DR   WPI; 1992-250033/30.
DR   N-PSDB; AAQ26509.
XX
PT   Acellular vaccine for immunisation against whooping cough -
PT   comprises protein uncontaminated by B. para:pertussis components
PT   and capable of binding antibodies which bind native P70 antigen
XX
PS   Claim 1; Fig 1; 20pp; English.
XX
CC   A cosmid library was constructed by transforming E.coli HB101 with
CC   recombinant cosmids prepared by partial digestion of B.parapertussis
CC   chromosomal DNA with Sau3A and cloning of 40-50kb fragments into the
CC   BamHI site of cosmid pHC79. The cosmids were screened with a 1.8kb
CC   ClaI fragment from the prn gene of B.pertussis. The insert from one
CC   positive colony, harbouring cosmid pBD811, was sequenced and found to
CC   contain an open reading frame encoding a 922 amino acid protein
CC   with calculated mol.wt. 95,177. This precursor protein ("P95") is
CC   processed in vivo to the P70 antigen of apparent mol. wt. 70,000 as
CC   determined by SDS-PAGE, but with actual mol.wt. 61kD. Antigenic
CC   fragments of the protein will be useful in developing an acellular
CC   vaccine against B.parapertussis. Preferred fragments include amino
CC   acids Pro577 to Pro612 or Ala574 to Pro612.
XX
SQ   Sequence   922 AA;

alignment_scores:
          Quality:  295.00          Length:   1083
            Ratio:  0.602             Gaps:     52
 Percent Similarity:  45.245    Percent Identity: 22.253

alignment_block:
 US-09-303-518D-649 x AAR25578   ..

 Align seg 1/1  to: AAR25578  from: 1  to: 922

  1417 GGGGAAAACCAAGGCTCGATCAGCGTGGGCGACGGTACAGTCATTTTGGA 1466
         ||||||:::|||   :::|||   :::
    45 GlyGluArgGlnHisGlyIleHisIle....................... 53

  1467 TCAGCAGGCAGACGATAAAGGCAAAAAACAAGCCTTTAGTGAAATCGGCT 1516
         :::|||:::|||        |||   :::   |||   :::
    54 .LysGlnSerAspGlyAlaGlyValArgThrAlaThrGlyThrThrIleL 70

  1517 TGGTCAGCGGCAGGGGTACGGTGCAACTGAATGCCGATAATCAGTTCAAC 1566
         ||||||||||||   :::      :::   :::|||
    70 ysValSerGlyArgGlnAlaGlnGlyValLeuLeuGluAsn......... 83

  1567 CCCGACAAACTCTATTTCGGCTTTCGCGGCGGACGTTTGGATTTAAACGG 1616
            :::   |||::::::|||   :::      :::||
    84 ...ProAlaAlaGluLeuArgPheGlnAsnGlySerValThrSerSerGl 99

  1617 GCATTCGCTTTCGTTCCACCGTATTCAAAATACCGATGAAGGGGCGATGA 1666
         |                     |||   |||||||||:::
    99 y......................GlnLeuPheAspGluGlyVal.... 106

  1667 TTGTCAACCACAATCAAGACAAAGAATCCACCGTTACCATTACAGGCAAT 1716
                      :::      :::|||||||||:::   ::::::
   107 ..............ArgArgPheLeuGlyThrValThrValLysAlaGly 118

  1717 AAA.........GATATTGCTACAACCGGCAATAACAACAGCTTGGATAG 1757
         |||         |||   ||||||   :::|||
   119 LysLeuValAlaAspHisAlaThrLeuAlaAsn................ 129

  1758 CAAAAAAGAAATTGCCTACAACGGTTGGTTTGGCGAGAAAGATACGACCA 1807
                                         :::||||||
   130 ................................ValSerAspThrArgA 135

  1808 AAACGAACGGGCGGCTCAACCTTGTT............TACCAGCCCGCC 1845
          :::|||         |||               |||   :::
   135 spAspAspGlyIleAlaLeuTyrValAlaGlyGluGlnAlaGlnAlaSer 151

  1846 GCAGAAGACCGCACCCTGCTGCTTTCCGGCGGAACAAATTTA......AA 1889
            |||   ||||||   :::||||||:::::::::      ::
   152 IleAlaAspSerThrLeuGlnGlyAlaGlyGlyValArgValGluArgGl 168

  1890 CGGCAACATCACGCAAACAAACGGCAAACTGTTTTCAGCGGC....... 1932
       :::::|||:::|||       ::::::   :::      :::|||
   168 yAlaAsnValThrValGlnArgSerThrIleValAspGlyGlyLeuHisI 185

  1933 ...........AGACCAACACCGCACGCCTACAATCATTTAAACGACCAT 1971
                  :::|||
   185 leGlyThrLeuGlnProLeu............................. 191

  1972 TGGTCGCAAAAAGAGGGCATTCCTCGCGGGGAAATCGTGTGGGACAACGA 2021
             |||   |||   :::|||   :::::::::|||      :::
   192 ......GlnProGluAspLeuProProSerArgValValLeuGlyAspTh 206

  2022 CTGGATCAACCGC..............................ACAT 2038
         ::::::                                  :::|
   206 rSerValThrAlaValProAlaSerGlyAlaProAlaAlaValPheValP 223

  2039 TTAAAGCGGAAAACTTCCAAATTAAAGGCGGACAGGCGGTGGTTTCCCGC 2088
        ||   |||:::::::::   :::   ||||||:::   :::   :::|||
   223 heGlyAlaAsnGluLeuThrValAspGlyGlyHisIleThrGlyGlyArg 239

  2089 AATGTTGCCAAAGTGAAAGGCGATTGGCATTTGAGCAATCACGCCCAAGC 2138
          ::::::                               :::   ||
   240 AlaAlaGly...............................ValAlaAl 245

  2139 AGTTTTTGGTGTCGCACCGCATCAAAGCCACACAATCTGTACACGTTCGG 2188
        |:::   |||:::      |||   :::::::::         |||:::|
   245 aMetAspGlyAlaIleValHisLeuGlnArgAlaThrIleArgArgGlyA 262

  2189 AC.............................................. 2190
        ||
   262 spAlaProAlaGlyGlyAlaValProGlyGlyAlaValProGlyGlyAla 278

  2191 ..................................TGGACGGGTCTGACAAA 2207
                                          |||   |||:::
   279 ValProGlyGlyPheGlyProLeuLeuAspGlyTrpTyrGlyVal..... 293

  2208 TTGTGTCGAAAAAACCATTACCGACGATAAAGTGATTGCTTCATTGACTA 2257
```

```
 293 ................................................... 293

2258 AGACCGACATCAGC...GGCAATGTCGATCTTGCCGATCACGCTCATTA 2304
         |||:::|||   :::::|||||||||||||:::
 294 .....AspValSerAspSerThrValAspLeuAlaGln............ 304

2305 AATCTCAGGGCTTGCCACACTCAACGGCAATCTTAGTGCAAATGGCGA 2354
     :::::::::                  |||:::::::   :::   |||:::
 305 SerIleValGluAlaProGlnLeuGlyAlaAlaIleArgAlaGlyArgGl 321

2355 TACACGTTATACAGTCAGCCACAAC......GCCACCCAAAACGGCAACC 2398
       ::::|||  |||||||||   :::       ::::::   :::||||||:
 321 yAlaArgValThrValSerGlyGlySerLeuSerAlaProHisGlyAsnV 338

2399 TTAGCCTCGTGGGCAATGCCCAAGCAACATTTAAT............... 2433
      ::      :::|||:::::            |||
 338 alIleGluThrGlyGlyGlyAlaArgArgPheProProProAlaSerPro 354

2434 ...CAAGCCACATTAAACGGCAACACATCGGCTTCGGGCAATGCT...TC 2477
       :::      ||||||:::::::::::    |||:::|||:::|||
 355 LeuSerIleThrLeuGlnAlaGlyAlaArgAlaGlnGlyArgAlaLeuLe 371

2478 ATTTAATCTAAGCGACCACGCCGTACAAAACGGCAGTCTGACGCTTTCCG 2527
       :::::::::      :::    |||:::           ||||||||||:::|
 371 uTyrArgValLeuProGluProValLys.........LeuThrLeuAlaG 385

2528 GCAACGCTAAGGCAAACGTAAGCCATTCCGCACTCAACGGTAATGTCTCC 2577
     |||:::|||::::::                        :::|||::::::
 385 lyGlyAlaGlnGly....................GlnGlyAspIleVal 394

2578 CTAGCCGATAAGGCAGTATTCCATTTGAAAGCAGCCGCTTTACCGGACA 2627
        :::                      |||           :::    |||
 395 AlaThr....................GluLeuProProIleProGlyAl 404

2628 AATCAGCGGCGGCAAGGATACGGCATTACACTTAAAAGACAGCGAATGGA 2677
          ||||||         |||:::||||||           ::::::::::::||||
 404 aSerSerGlyProLeuAspValAlaLeu...AlaSerGlnAlaArgTrpT 420

2678 CGCTGCCGTCAGGCACGGAATTAGGCAATTTAAACCTTGACAACGCCACC 2727
      ||       :::            :::    :::|||::::::||||||||||||
 420 hr...GlyAlaThrArgAlaValAspSerLeuSerIleAspAsnAlaThr 435

2728 ATTACACTCAATTCCGCCTATCGCCACGATGCGGCAGGGGCGCAAACCGG 2777
        ::::::
 436 TrpValMet......................................... 438

2778 CAGTGCGACAGATGCGCCGCGCCGCCGTTCGCGCCGTTCGCGCCGTTCCC 2827
             ||||||                        ::::|||:::    :::
 439 .......ThrAsp...................AsnSerAsnValGlyA 446

2828 TATTATCCGTTACACCGCCAACTTCGGTAGAA.................. 2859
       |||    ::::::          ||||||:::
 446 laLeuArgLeuAlaSerAspGlySerValAspPheGlnGlnProAlaGlu 462

2860 ...TCCCGTTTCAACACGCTGACGGTAAACGGCAAATTGAACGGTCAGGG 2906
       :::||||||:::::::|||   |||:::        |||    |||:::||
 463 AlaGlyArgPheLysValLeuMetValAsp...ThrLeuAlaGlySerGl 478

2907 AACATTCCGCTTTATGTCGGAACTCTTCGGCTACCGCAGCGACAAATTGA 2956
     |    |||||||:::                        ||||||||||||
 478 yLeuPheArgMetAsnValPheAlaAspLeuGlyLeuSerAspLysLeuV 495

2957 AGCTGGCGGAAAGTTCCGAAGGCACTTACACCTTGGCGGTCAACAATACC 3006
       :::    ::::::::::::|||     :::    |||     |||:::|||:::
 495 alValMetArgAspAlaSerGlyGlnHisArgLeuTrpValArgAsnSer 511

3007 GGCAACGAACCTGCAAGCCTCGAACAATTGACGGTAGTGGAAGGAAAAGA 3056
     |||:::||||||||||||||     :::     :::     :::|||:::
 512 GlySerGluProAlaSerGlyAsnThrMetLeuLeuValGln........ 525

3057 CAACAAACCGCTGTCCGAAAACCTTAATTTCACCCTGCAAAACGAA.... 3102
             |||      ::::::          :::|||||||||     |||:::
 526 ....ThrProArgGlySerAlaAlaThrPheThrLeuAlaAsnLysAspG 541
```

```
3103 ..CACGTCGATGCCGGCGCGTGGCGTTACCAACTCATCCGCAAAGACGGC 3150
         ||||||    |||::::::|||||||:::|||
 541 lyLysValAspIleGlyThrTyrArgTyrArgLeu............... 552

3151 GAGTTCCGCCTGCATAATCCGGTCAAAGAACAAGAGCTTTCCGACAAACT 3200

 552 ................................................... 552

3201 CGGCAAGGCAGAAGCCAAAAAACAGGCGGAAAAAGACAACGCGCAAAGCC 3250
                             |||    :::    |||:::|||
 553 ........................AlaAlaAsnGlyAsnGlyGlnTrp. 560

3251 TTGACGCGCTGATTGCGGCCGGGCGCGATGCCGTCGAAAAGACAGAAAGC 3300
        :::|||:::::::|||     |||     |||:::          |||
 561 .....SerLeuValGlyAla.LysAlaProProAla......ProLysPr 573

3301 GTTGCCGAACCGGCCCGGCAGGCAGGCGGGGAAAATGTCGGCATTATGCA 3350
        |||     :::||||||    |||    |||     :::
 573 oAlaPro...GlnProGlyProGlnProGlyProGln............. 584

3351 GGCGGAGGAAGAGAAAAAACGGGTGCAGGCGGATAAAGACACCGCCTTGG 3400
                                                  ||||||
 585 .........................................ProPro... 586

3401 CGAAACAGCGCGAAGCGGAAACCCGGCCGGCTACCACCGCCTTCCCCCGC 3450
                                         ||||||    ||||||:
 587 .............................GlnProProGln.ProProG 593

3451 GCCCGCCGCGCCCGCCGGGATTTGCCGCAACTGCAACCCCAA........ 3492
      ::||||||:::||||||  :::      ||||||    ||||||:::
 593 lnProProGlnProPro.GlnProProGlnArgGlnProGluAlaProAl 609

3493 .CCGCAGCCCCAACCGCAGCGCGACCTGATCAGCCGTTATGCCAATAGCG 3541
      |||||||||                 |||:::|||    |||       ||||||:::
 609 aProGlnProProAlaGlyArgGluLeu...SerAlaAlaAlaAsnAlaA 625

3542 GTTTGAGTGAATTTTCCGCCACGCTCAACAGCGTT......TTCGCCGTA 3585
      ::::::::          ::::::    |||    |||:::          :::|||
 625 laValAsnThrGlyGlyValGlyLeuAlaSerThrLeuTrpTyrAlaGlu 641

3586 CAGGACGAATTAGACCGCGTATTTGCCGAAGACCGCCGCAAC........ 3627
      ::::::    |||:::::::    ::::::|||    |||    |||
 642 SerAsnAlaLeuSerLysArgLeuGlyGluLeuArgLeuAsnProAspAl 658

3628 ....GCCGTTTGGACAAGCGGCATCCGGGACACCAAACACTACCGTTCGC 3673
        ::::::|||         |||:::    :::    ::::::         ::::
 658 aGlyGlyAlaTrpGlyArgGlyPheAlaGlnArgGlnGlnLeuAspAsnA 675

3674 AAGATTTCCGCGCCTACCGCCAACAAACCGACCTGCGCCAAATCGGTATG 3723
      ::       |||    :::    |||::::::          ::::::|||
 675 rgAlaGlyArgArgPheAspGlnLysValAlaGlyPheGluLeuGlyAla 691

3724 CAGAAAAACCTC............GGCAGCGGGCGCGTCGGCATCCTGTT 3761
      :::      :::            |||        :::::::|||    |||
 692 AspHisAlaValAlaValAlaGlyGlyArgTrpHisLeuGlyGlyLeuAl 708

3762 TTCGCACAACCGGACCGAAAACACCTTC...GACGACGGCATCGGCAACT 3808
       :::::::::|||    ::::::    |||          ||||||    |||::::
 708 aGlyTyrThrArgGlyAspArgGlyPheThrGlyAspGlyGlyGlyHisT 725

3809 CGGCACGGCTTGCCCACGGCGCCGTTTTCGGGCAATACGGCATCGACAGG 3858
      ::       :::         |||:::                :::    :::
 725 hrAspSerValHisValGlyGlyTyrAlaThrTyrIleAlaAsnSerGly 741

3859 TTCTACATCGGCATCAGCGCGGGCGCGGGTTTTAGCAGCGGCAGCCTT.. 3906
     |||||||:::      :::         |||:::          :::::::::::
 742 PheTyrLeuAspAlaThrLeuArgAlaSerArgLeuGluAsnAspPheLy 758

3907 ..........TCAGACGGC......ATCGGAGGCAAAATCCGCCGCCGCG 3940
                ||||||||         :::    ||||||    |||       :::
 758 sValAlaGlySerAspGlyTyrAlaValLysGlyLysTyrArgThrHisG 775
```

```
3941 TGCTGCATTACGGCATTCAGGCACGATACCGC...GCCGGTTTCGGCGGA 3987
       :::      ::::::::|||      |||    |||       |||
 775 lyValGlyValSerLeuGluAlaGlyArgArgPheAlaHisAlaAspGly 791

3988 TTCGGCATCGAACCGCACATCGGCGCAACGCGCTATTTCGTCCAAAAAGC 4037
     :::    :::||||||:::          :::   :::   |||       ::
 792 TrpPheLeuGluProGlnAlaGluLeuAlaValPheArgValGlyGlyGl 808

4038 GGATTACCGCTACGAAAAC...GTCAATATCGCCACCCCCGGCCTTGCAT 4084
      :   ||||||      |||   :::::::::          |||    :::
 808 yAlaTyrArgAlaAlaAsnGlyLeuArgValArgAspGluGlyGlySerS 825

4085 TCAACCGCTACCGCGCGGGCATTAAGGCAGATTATTCATTCAAACCGGCG 4134
                  |||    |||:::::::::          ::::::   |||
 825 erValLeuGlyArgLeuGlyLeuGluValGlyLysArgIleGluLeuAla 841

4135 CAACACATTTCCATCACGCCTTATTTGAGCCTGTCCTATACCGATGCCGC 4184
                 ::::::   ||||||::::::   |||        :::
 842 GlyGlyArgGlnValGlnProTyrIleLysAlaSerValLeuGlnGluPh 858

4185 TTCGGGCAAAGTCCGAACACGCGTCAATACCGCCGTATTGGCTCAGGATT 4234
      :::|||            :::|||:::|||         :::          ::::
 858 eAspGlyAlaGlyThrValArgThrAsnGlyIleAlaHisArgThrGluL 875

4235 TCGGCAAAACCCGCAGTGCGGAATGGGGCGTAAACGCCGAAATCAAAGGT 4284
      ::       ||||||   ||||||   |||
 875 euArgGlyThrArg...AlaGluLeuGly..................... 883

4285 TTCACGCTGTCCCTCCACGCTGCCGCCGCCAAAGGCCCGCAACTGGAAGC 4334
               |||::::::   ||||||   ::::::|||    :::|||    ||
 884 ......LeuGlyMetAlaAlaAlaLeuGlyArgGlyHisSerLeuTyrAl 898

4335 G......CAACACAGCGCGGGCATCAAATTAGGCTACCGCTGG 4371
      |       :::::::|||   |||   ||||||:::       |||
 898 aSerTyrGluTyrSerLysGlyProLysLeuAlaMetProTrp 912
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:AAR41732

seq_documentation_block:

```
ID   AAR41732 standard; Protein; 1529 AA.
XX
AC   AAR41732;
XX
DT   26-APR-1994  (first entry)
XX
DE   High molecular weight protein 4 (HMW4).
XX
KW   HMW; high molecular weight protein; virus; vaccine; influenza;
KW   epitope; immunity; haemophilus influenzae.
XX
OS   Haemophilus influenzae.
XX
PN   WO9319090-A.
XX
PD   30-SEP-1993.
XX
PF   16-MAR-1993;    93WO-US02166.
XX
PR   16-MAR-1992;    92GB-0005704.
XX
PA   (BARE/) BARENKAMP S J.
PA   (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
XX
PI   Barenkamp SJ;
XX
DR   WPI; 1993-320683/40.
DR   N-PSDB; AAQ49511.
XX
PT   High molecular weight surface proteins - of non-typeable
PT   haemophilus which exhibit immunogenic properties
XX
PS   Claim 6; Figure 10; 100pp; English.
XX
CC   The isolation and purification of the high molecular weight protein
CC   enables the identification of the major protective epitopes of the
CC   protein by conventional epitope mapping.  These epitopes can then be
CC   synthesised using standard techniques and incorporated into fully
CC   synthetic or recombinant vaccines.
XX
SQ   Sequence   1529 AA;
```

```
alignment_scores:
        Quality:  289.50          Length:   1196
          Ratio:  0.497             Gaps:     61
Percent Similarity:  48.746   Percent Identity:  20.652

alignment_block:
 US-09-303-518D-649 x AAR41732   ..

 Align seg 1/1  to: AAR41732  from: 1  to: 1529

 289 TCGCGTAACGGCGTGGCGGCATTGGTGGGCGATCAATATATTGTGAGCGT 338
     :::||||||::::::    |||    :::       |||      ::::::||
 507 AlaArgAsnAlaThrAspAlaGlnIleThrAlaGlnGlyThrIleThrVa 523

 339 GGCACATAACGGC.........GGCTATAACAACGTTGATTTTGGTGCGG 379
     |     :::                ::::|||||||||::::::::::::
 523 lAsnLysAspAspLysGlnPheArgPheAsnAsnValSerIleAsnGlyT 540

 380 AAGGAAGAAATCCCGATCAACATCGTTTTACTTATAAAATTGTGAAACGG 429
       |||::::::           ::::|||         |||:::::::::
 540 hrGlyLysGly.........LeuLysPhe.........IleAlaAsnGln 550

 430 AATAATTATAAAGCAGGGACTAAAGGCCATCCTTATGGCGGCGATTATCA 479
     ||||||:::                  |||    :::   |||:::    ::
 551 AsnAsnPheThr...............HisLysPheAspGlyGluIleAs 562

 480 TATGCCGCGTTTGCATAAATTTGTCACAGATGCAGAACCTGTTGAAATGA 529
     :
 562 n................................................. 562

 530 CCAGTTATATGGATGGGCGGAAATATATCGATCAAAATAATTACCCTGAC 579
             :::::::|||         |||:::|||::::::       |||
 563 ........IleSerGlyIleValThrIleAsnGlnThrThrLysLysAsp 576

 580 CGTGTTCGTATTGGGGCAGGCAGGCAATATTGGCGATCTGATGAAGATGA 629
                          :::||||||::::::::::::::|||::
 577 ....................ValLysTyrTrpAsnAlaSerLysAspSe 586

 630 GCCCAATAACCGCGAAAGT...TCATATCATATTGCAAGTGCGTATTCTT 676
     :     |||    :::|||    :::    :::    ::::::   :::::::
 586 rTyrTrpAsnValSerSerLeuThrLeuAsnThrValGlnLysPheThrP 603

 677 GGCTC...........GTTGGTGGCAAT...................... 693
     :::::                 |||:::|||
 603 heIleLysPheValAspSerGlySerAsnSerGlnAspLeuArgSerSer 619

 694 ......ACCTTTGCACAA.........AATGGATCAGGTGGTGGCACA.. 726
           :::||||||              ||||||    ||||||   |||
 620 ArgArgSerPheAlaGlyValHisPheAsnGlyIleGlyGlyLysThrAs 636

 727 .GTCAACTTAGGTAGTGAA..............AAAATTAAACATAGC. 759
         |||:::|||::::::                  |||:::|||     :::
 636 nPheAsnIleGlyAlaAsnAlaLysAlaLeuPheLysLeuLysProAsnA 653

 760 ..........CCATATGGTTTTTTACCA...................... 777
                 |||         ||||||
 653 laAlaThrAspProLysLysGluLeuProIleThrPheAsnAlaAsnIle 669

 778 ......ACAGGAGGCTCATTTGGCGACAGTGGCTCACCAATGTTTATCTA 821
           ||||||:::|||            ::::::|||    ||||||
 670 ThrAlaThrGlyAsnSer.........AspSerSerValMetPheAspIl 683

 822 TGATGCCCAAAAGCAAAAGTGGTTA...........ATTAATGGGGTAT 859
         |||:::       :::                   ::::::::::::
```

```
 683 eHisAlaAsnLeuThrSerArgAlaAlaGlyIleAsnMetAspSerIleA 700

 860 TGCAAACGGGC..................AACCCCTATATAGGAAAAAGC 891
          ||||||                  :::   :::     :::|||
 700 snIleThrGlyGlyLeuAspPheSerIleThrSerHisAsnArgAsnSer 716

 892 AATGGCTTCCAGCTGGTTCGT............................. 912
     |||:::|||:::::   :::
 717 AsnAlaPheGluIleLysLysAspLeuThrIleAsnAlaThrGlySerAs 733

 913 ...................AAAGATTGGTCTATGATGAAATCTTTGCTG 943
                        ||||||   ||||||:::|||
 733 nPheSerLeuLysGlnThrLysAspSerPheTyrAsnGluTyrSerLysH 750

 944 GAGATACCCATTCAGTATTCTACGAACCACGTCAAAATGGGAAATACTCT 993
              :::|||                      :::|||:::   :::
 750 isAlaIleAsnSerSerHisAsnLeuThrIleLeuGlyGlyAsnValThr 766

 994 TTTAACGACGATAATAATGGC.........ACAGGAAAAATCAATGCCAA 1034
     ::::::   :::|||::::::         ||||||:::||||||
 767 LeuGlyGlyGluAsnSerSerSerSerIleThrGlyAsnIleAsnIleTh 783

1035 ACATGAACACAAT.........................TCTCTGCCTAATA 1060
      ::::::   |||                         :::     :::
 783 rAsnLysAlaAsnValThrLeuGlnAlaAspThrSerAsnSerAsnThrG 800

1061 GATTAAAAACACGAACCGTTCAATTGTTTAATGTTTCTTTATCCGAGACA 1110
        ||||||   |||||||:::   |||   |||:::|||:::
 800 lyLeuLysLysArgThrLeuThrLeuGlyAsnIleSerVal......... 813

1111 GCAAGAGAACCTGTTTATCATGCTGCAGGTGGTGTCAACAGTTATCGACC 1160
           |||               :::|||:::     :::
 814 ......GluGlyAsnLeuSerLeuThrGlyAlaAsnAlaAsnIleValGl 828

1161 CAGACTGAATAATGGAGAAAATATTTCCTTTATTGACGAAGGAAAAGGCG 1210
       :::|||:::   :::|||:::   :::|||      |||::::::   :
 828 yAsnLeuSerIleAlaGluAspSerThrPheLysGlyGluAlaSerAspA 845

1211 AATTGATACTT.........................ACCAGCAACATCAAT 1236
      ::|||   :::                        |||:::|||||||||
 845 snLeuAsnIleThrGlyThrPheThrAsnAsnGlyThrAlaAsnIleAsn 861

1237 CAAGGTGCTGGAGGATTATATTTCCAAGGAGATTTACGGTCTCGCCTGA 1286
           |||   :::   :::|||||||||:::
 862 IleLysGlnGlyValValLysLeuGlnGlyAspIleIle........... 874

1287 AAATAACGAAACTTGGCAAGGCGCGGGCGTTCATATCAGTGAAGACAGTA 1336
                       :::   :::|||::::::|||:::   ::::::
 875 ................AsnLysGlyGlyLeuAsnIleThrThrAsnAlaS 886

1337 CCGTTACTTGGAAA......GTAAACGGC...GTGGCAAACGACCGCCTG 1377
      ::    |||   |||        :::||||||   ::::::|||:::
 886 erGlyThrGlnLysThrIleIleAsnGlyAsnIleThrAsnGlu...... 900

1378 TCCAAAATCGGCAAAGGCACGCTG.........CACGTTCAAGCCAAAGG 1418
                 ||||||   |||            :::::::::|||   ::
 901 ...........LysGlyAspLeuAsnIleLysAsnIleLysAlaAspAl 913

1419 GGAAAACCAA......GGCTCGATCAGCGTGGGCGACGGTACAGTCATTT 1462
      :|||   |||         |||:::||||||     :::|||::::::   :
 913 aGluIleGlnIleGlyGlyAsnIleSerGlnLysGluGlyAsnLeuThrI 930

1463 TG................................................ 1464
     ::
 930 leSerSerAspLysValAsnIleThrAsnGlnIleThrIleLysAlaGly 946

1465 .....................GATCAGCAGGCAGACGATAAAGGCAAAAA 1493
                          :::::::::|||::::::   :::
 947 ValGluGlyGlyArgSerAspSerSerGluAlaGluAsnAlaAsnLeuTh 963

1494 ACAAGCCTTTAGTGAAATCGGCTTG...............GTCAGCGGCA 1528
              :::|||:::   |||                :::||||||
 963 rIleGlnThrLysGluLeuLysLeuAlaGlyAspLeuAsnIleSerGlyP 980

1529 GGGGTACGGTGCAACTGAATGCCGATAAT..................... 1557
       :::   ::::::::::::|||   |||
 980 heAsnLysAlaGluIleThrAlaLysAsnGlySerAspLeuThrIleGly 996

1558 ...............CAGTTCAACCCCGACAAACTCTATTTCGGC...TT 1589
                    :::    :::        |||:::    |||
 997 AsnAlaSerGlyGlyAsnAlaAspAlaLysLysValThrPheAspLysVa 1013

1590 TCGCGGCGGACGTTTGGATTTAAACGGGCATTCGCTTTCGTTCCACCGTA 1639
      :::   ::::::::::::   :::||||||:::::::::::::::
1013 lLysAspSerLysIleSerThrAspGlyHisAsnValThrLeuAsnSerG 1030

1640 TTCAAAATACCGATGAAGGGGCGATGATTGTCAACCACAATCAAGACAAA 1689
         :::|||:::::::|||:::           |||    :::::|||:::
1030 luValLysThrSerAsnGlySerSer......AsnAlaGlyAsnAspAsn 1044

1690 GAATCCACCGTTACCATTACAGGCAATAAAGATATTGCTACAACCGGCAA 1739
     ::::::   :::||||||:::::   ||||||:::::::::
1045 SerThrGlyLeuThrIleSerAla...LysAspValThrVal........ 1057

1740 TAACAACAGCTTGGATAGCAAAAAAGAAATTGCCTACAACGGTTGGTTTG 1789
      ||||||::::::   |||   |||   |||      ::::::      |
1058 .AsnAsnAsnValThrSerHisLysThrIleAsnIleSerAlaAlaAlaG 1074

1790 GCGAGAAAGATACGACCAAAACGAACGGGCGGCTCAACCTTGTTTACCAG 1839
     ||:::       |||||||||   :::      :::|||
1074 lyAsnVal...ThrThrLysGluGlyThrThrIleAsn............ 1085

1840 CCCGCCGCAGAAGACCGCACCCTGCTGCTTTCCGGCGGAACAAATTTAAA 1889
        |||:::         :::   :::   :::   :::   ::::::::
1086 ...AlaThrThrGlySerValGluValThrAlaGlnAsnGlyThrIleLy 1101

1890 CGGCAACATCACGCAAACAAACGGCAAACTGTTTTTCAGCGGCAGACCAA 1939
     :||||||||||||:::   |||                          |
1101 sGlyAsnIleThrSerGlnAsn.......................ValT 1110

1940 CACCGCACGCCTACAATCATTTAAACGACCATTGGTCGCAAAAAGAGGGC 1989
     ||      |||     ::::::|||
1110 hrValThrAlaThrGluAsnLeu........................... 1117

1990 ATTCCTCGCGGGGAAATCGTGTGGGACAACGACTGGATCAACCGCACATT 2039
                    ::::::   :::|||      ||||||   |||
1118 ...............ValThrThrGluAsnAlaValIleAsnAlaThrSe 1129

2040 TAAAGCGGAAAACTTCCAAATTAAAGGCGGACAGGCGGTGGTTTCCCGCA 2089
        :::   |||::::::   |||   |||:::      ::::::   :
1129 rGlyThrValAsnIleSerThrLysThrGlyAsp......IleLysGlyG 1144

2090 ATGTTGCCAAAGTGAAAGGCGATTGGCATTTGAGCAATCACGCCCAAGCA 2139
     :::::   :::::::::|||:::   :::::::::         :::::::::
1144 lyIleGluSerThrSerGlyAsnValAsnIleThrAlaSerGlyAsnThr 1160

2140 GTTTTTGGTGTCGCACCGCATCAAAGCCACACAATCTGTACACGTTCGGA 2189
     :::      |||:::        ::::::   :::   :::   :::||
1161 Leu...LysValSerAsnIleThrGlyGlnAspValThrValThrAlaAs 1176

2190 CTGGACGGGTCTGACAAATTGTGTCGAAAAAACCATTACCGACGATAAAG 2239
     |         :::|||||||:::   :::   :::||||||:::
1176 pAlaGlyAlaLeuThrThrThrAlaGlySerThrIleSerAlaThrThrG 1193

2240 TGATTGCTTCATTG...ACTAAGACC...GACATCAGCGGCAATGTCGAT 2283
        |||::::::   |||||||||   ||||||:::|||:::|||:::
1193 lyAsnAlaAsnIleThrThrLysThrGlyAspIleAsnGlyLysValGlu 1209

2284 CTTGCCGAT...CACGCTCATTTAAATCTCACAGGGCTTGCCACACTCAA 2330
        ::::::      :::   |||      ||||||   :::
1210 SerSerSerGlySerValThrLeuValAlaThrGlyAlaThrLeuAlaVa 1226

2331 CGGCAATCTTAGTGCAAATGGCGATACA..................CGTT 2362
      ||||||:::|||:::|||      |||                   :::
1226 lGlyAsnIleSerGlyAsnThrValThrIleThrAlaAspSerGlyLysL 1243
```

```
2363 ATACAGTCAGCCACAACGCCACCCAAAACGGCAACCTTAGCCTCGTGGGC 2412
       |||   :::   ::::::|||  ||||||:::   |||:::::
1243 euThrSerThrValGlySerThrIleAsnGlyThrAsnSerValThrThr 1259

2413 AATGCCCAAGCAACATTTAATCAAGCCACATTAAACGGCAACACATCGGC 2462
     ::::::|||:::          ::::::|||::::::||||||||
1260 SerSerGlnSerGlyAspIleGluGlyThrIleSerGlyAsnThr..... 1274

2463 TTCGGGCAATGCTTCATTTAATCTAAGCGACCACGCCGTACAAAACGGCA 2512
                       |||:::::        ::::::::|||:
1275 ................ValAsnValThr.........AlaSerThrGlyA 1283

2513 GTCTGACGCTTTCCGGCAACGCTAAGGCAAACGTAAGCCATTCCGCACTC 2562
     ::||||||::::::::::::||||||::::::::::::::::::|||
1283 spLeuThrIleGlyAsnSerAlaLysValGluAlaLysAsnGlyAla... 1298

2563 AACGGTAATGTCTCCCTAGCCGATAAGGCAGTATTCCATTTTGAAAGCAG 2612
                                |||::::::     ||||||::
1299 .........................AlaThrLeuThrAlaGluSerGl 1306

2613 CCGCTTTACCGGACAA...............ATCAGCGGCGGCAAGGATA 2647
     :::::::|||   |||                  |||::::::   :::|
1306 yLysLeuThrThrGlnThrGlySerSerIleThrSerSerAsnGlyGlnT 1323

2648 CGGCATTACACTTAAAAGACAGCGAATGGACGCTGCCGTCAGGCACGGAA 2697
     ||:::|||    |||||||||                    :::
1323 hrThrLeuThrAlaLysAspSer.....................SerIle 1332

2698 TTAGGCAATTTAAACCTTGACAACGCCACCATTACACTCAATTCCGCCTA 2747
       ||||||:::|||          |||::::::||||||||||:::
1333 AlaGlyAsnIleAsnAla......AlaAsnValThrLeuAsnThr..... 1345

2748 TCGCCACGATGCGGCAGGGGCGCAAACCGGCAGTGCGACAGATGCGCCGC 2797
                             ||||||:::
1346 ........................ThrGlyThr............... 1348

2798 GCCGCCGTTCGCGCCGTTCGCGCCGTTCCCTATTATCCGTTACACCGCCA 2847
                                     |||::::::|||
1349 ................................LeuThrThrThrGlyAsp 1354

2848 ACTTCGGTAGAATCCCGTTTCAACACGCTGACGGTAAACGGCAAATTGAA 2897
     :::::::::::::::        :::|||||||||:::|||:::|||
1355 SerLysIleAsnAlaThrSerGlyThrLeuThrIleAsnAlaLys..... 1369

2898 CGGTCAGGGAACATTCCGCTTTATGTCGGAACTCTTCGGCTACCGCAGCG 2947
                                                 :::
1370 ..........................................AspA 1371

2948 ACAAATTGAAGCTGGCGGAAAGTTCCGAAGGCACTTACACCTTGGCGGTC 2997
        ||||||      |||    |||:::::   |||    :::   |||:::
1371 laLysLeuAspGlyAlaAlaSerGlyAspArgThrValValAsnAlaThr 1387

2998 AACAATACCGGCAACGAACCTGCAAGCCTCGAACAATTGACGGTAGTGGA 3047
     |||   :::|||:::
1388 AsnAlaSerGlySer.................................. 1392

3048 AGGAAAAGACAACAAACCGCTGTCCGAAAACCTTAATTTCACCCTGCAAA 3097
       |||:::   :::           |||:::::::::|||:::|||    :::
1393 .GlyAsnValThrAlaLysThrSerSerSerValAsnIleThrGlyAspL 1409

3098 ACGAACACGTCGATGCCGGCGCGTGGCGTTACCAACTCATCCGCAAAGAC 3147
        :::   :::::::::                 ::::::|||   ::::::
1409 euAsnThrIleAsnGly...........LeuAsnIleIleSerGluAsn 1421

3148 GGCGAGTTCCGCCTGCATAATCCGGTCAAAGAACAAGAGCTTTCCGACAA 3197
      |||            :::|||   |||                    ::
1422 Gly...........ArgAsnThrVal.....................Ar 1427

3198 ACTCGGCAAGGCAGAAGCCAAAAAACAGGCGGAAAAAGACAACGCGCAAA 3247
      :|||         |||        :::       :::   ::::::    |
1427 gLeuArgGlyLysGluIleAspValLysTyrIleGlnProGlyValAlaS 1444

3248 GCCTTGACGCGCTGATTGCGGCCGGGCGCGATGCCGTCGAAAAGACAGAA 3297

      ||::::::     :::|||   |||   |||     ::::::||||||::::::
1444 erValGluGluValIleGluAlaLysArg...ValLeuGluLysValLys 1459

3298 AGCGTTGCCGAACCGGCCCGGCAGGCAGGCGGGGAAAATGTCGGCATTAT 3347
     ::::::::::::             |||:::::::   ::::::   :::|||:::
1460 AspLeuSerAspGluGluArgGluThrLeuAlaLys...LeuGlyValSe 1475

3348 GCAGGCGGAGGAAGAGAAAAAACGGGTGCAGGCGGATAAAGACACCGCCT 3397
                     :::   |||:::   ::::::         ::::
1475 rAlaVal...........ArgPheValGluProAsnAsnAlaIleThrV 1488

3398 TGGCGAAACAGCGCGAAGCGGAAACCCGGCCGGCTACC 3435
     ::        |||:::|||      |||:::|||::::::
1488 alAsnThrGlnAsnGluPheThrThrLysProSerSer 1500
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:AAW30292

```
seq_documentation_block:
ID   AAW30292 standard; Protein; 1601 AA.
XX
AC   AAW30292;
XX
DT   14-APR-1998  (first entry)
XX
DE   Non-typeable Haemophilus high mol.wt. surface protein HMW4.
XX
KW   Non-typeable Haemophilus; high molecular weight surface protein;
KW   HMW4; immunogen; vaccine; otitis media.
XX
OS   Haemophilus influenzae strain 5.
XX
FH   Key                Location/Qualifiers
FT   Misc-difference    372
FT                      /note= "encoded by TCT"
FT   Misc-difference    400
FT                      /note= "encoded by AAT"
XX
PN   WO9736914-A1.
XX
PD   09-OCT-1997.
XX
PF   01-APR-1997;   97WO-US04707.
XX
PR   01-APR-1996;   96US-0617697.
XX
PA   (BARE/) BARENKAMP S J.
XX
PI   Barenkamp SJ;
XX
DR   WPI; 1997-503038/46.
DR   N-PSDB; AAT90993.
XX
PT   High molecular weight proteins of non-typeable Haemophilus
PT   influenzae - useful for vaccine production
XX
PS   Claim 1; Page 97-102; 183pp; English.
XX
CC   This protein comprises the high molecular weight surface protein
CC   HMW4 (123 kDa) of non-typeable Haemophilus influenzae strain 5 that
CC   has the immunological ability to protect against disease caused by
CC   a non-typeable Haemophilus strain and is characterised by at least
CC   one surface-exposed B-cell epitope that is recognised by monoclonal
CC   antibody AD6.  The HMW4 amino acid sequence was deduced from an
CC   isolated hmw4 gene (see AAT90993).  HMW1 (see AAW30293), HMW2 (see
CC   AAW30294) and HMW3 (see AAW30291) have also been identified.  A
CC   conjugate comprising HMW4 linked to an antigen, hapten or
CC   polysaccharide, and a synthetic peptide of 6-150 amino acids
CC   corresponding to at least protective epitope of HMW4 are also
CC   claimed.  HMW proteins, conjugates and peptides can be used in
CC   vaccines, as immunogens for preparation of antibodies and as
CC   antigens for detection of these antibodies.
XX
SQ   Sequence   1601 AA;
```

```
alignment_scores:
         Quality:  286.50          Length:   1189
           Ratio:  0.491             Gaps:     64
 Percent Similarity:  49.117   Percent Identity:  21.026

alignment_block:
 US-09-303-518D-649 x AAW30292   ..

 Align seg 1/1  to: AAW30292  from: 1  to: 1601

    289 TCGCGTAACGGCGTGGCGGCATTGGTGGGCGATCAATATATTGTGAGCGT 338
         :::|||||:::::   |||   :::      |||     ::::::||
    579 AlaArgAsnAlaThrAspAlaGlnIleThrAlaGlnGlyThrIleThrVa 595

    339 GGCACATAACGGC.........GGCTATAACAACGTTGATTTTGGTGCGG 379
         |      :::        ::::|||||||||:::::::::::
    595 lAsnLysAspAspLysGlnPheArgPheAsnAsnValSerIleAsnGlyT 612

    380 AAGGAAGAAATCCCGATCAACATCGTTTTACTTATAAAATTGTGAAACGG 429
          |||:::::         ::::|||         |||::::::::
    612 hrGlyLysGly.........LeuLysPhe.........IleAlaAsnGln 622

    430 AATAATTATAAAGCAGGGACTAAAGGCCATCCTTATGGCGGCGATTATCA 479
        ||||||:::                   |||   :::   |||:::   ::
    623 AsnAsnPheThr...............HisLysPheAspGlyGluIleAs 634

    480 TATGCCGCGTTTGCATAAATTTGTCACAGATGCAGAACCTGTTGAAATGA 529
        :
    634 n................................................. 634

    530 CCAGTTATATGGATGGGCGGAAATATATCGATCAAAATAATTACCCTGAC 579
                :::::|||          |||:::|||::::::    |||
    635 ........IleSerGlyIleValThrIleAsnGlnThrThrLysLysAsp 648

    580 CGTGTTCGTATTGGGGCAGGCAGGCAATATTGGCGATCTGATGAAGATGA 629
                             :::||||||::::::::::::|||:::
    649 ....................ValLysTyrTrpAsnAlaSerLysAspSe 658

    630 GCCCAATAACCGCGAAAGT...TCATATCATATTGCAAGTGCGTATTCTT 676
        :       |||   :::|||   :::   :::   ::::::   :::::::
    658 rTyrTrpAsnValSerSerLeuThrLeuAsnThrValGlnLysPheThrP 675

    677 GGCTC...........GTTGGTGGCAAT...................... 693
        :::::          |||::::|||
    675 heIleLysPheValAspSerGlySerAsnSerGlnAspLeuArgSerSer 691

    694 ......ACCTTTGCACAA.........AATGGATCAGGTGGTGGCACA.. 726
              :::||||||          ||||||   ||||||   |||
    692 ArgArgSerPheAlaGlyValHisPheAsnGlyIleGlyGlyLysThrAs 708

    727 .GTCAACTTAGGTAGTGAA..............AAAATTAAACATAGC. 759
           |||:::|||:::::::              |||::::|||   :::
    708 nPheAsnIleGlyAlaAsnAlaLysAlaLeuPheLysLeuLysProAsnA 725

    760 ..........CCATATGGTTTTTTACCA...................... 777
                    |||     ||||||
    725 laAlaThrAspProLysLysGluLeuProIleThrPheAsnAlaAsnIle 741

    778 ......ACAGGAGGCTCATTTGGCGACAGTGGCTCACCAATGTTTATCTA 821
              ||||||:::|||          :::::::|||   ||||||
    742 ThrAlaThrGlyAsnSer.........AspSerSerValMetPheAspIl 755

    822 TGATGCCCAAAAGCAAAAGTGGTTA...........ATTAATGGGGTAT 859
           |||:::       :::               ::::::::::::
    755 eHisAlaAsnLeuThrSerArgAlaAlaGlyIleAsnMetAspSerIleA 772

    860 TGCAAACGGGC.................AACCCCTATATAGGAAAAAGC 891
          ||||||                   :::   :::      :::|||
    772 snIleThrGlyGlyLeuAspPheSerIleThrSerHisAsnArgAsnSer 788

    892 AATGGCTTCCAGCTGGTTCGT............................ 912

        |||:::|||::::::   :::
    789 AsnAlaPheGluIleLysLysAspLeuThrIleAsnAlaThrGlySerAs 805

    913 ...................AAAGATTGGTTCTATGATGAAATCTTTGCTG 943
                            ||||||   ||||||::::|||
    805 nPheSerLeuLysGlnThrLysAspSerPheTyrAsnGluTyrSerLysH 822

    944 GAGATACCCATTCAGTATTCTACGAACCACGTCAAAATGGGAAATACTCT 993
              :::|||                        :::|||:::   :::
    822 isAlaIleAsnSerSerHisAsnLeuThrIleLeuGlyGlyAsnValThr 838

    994 TTTAACGACGATAATAATGGC.........ACAGGAAAAATCAATGCCAA 1034
        ::::::   :::|||::::::          ||||||:::||||||
    839 LeuGlyGlyGluAsnSerSerSerSerIleThrGlyAsnIleAsnIleTh 855

   1035 ACATGAACACAAT.......................TCTCTGCCTAATA 1060
        ::::::   |||                          :::   :::
    855 rAsnLysAlaAsnValThrLeuGlnAlaAspThrSerAsnSerAsnThrG 872

   1061 GATTAAAAACACGAACCGTTCAATTGTTTAATGTTTCTTTATCCGAGACA 1110
          ||||||   |||||||:::   |||   |||:::|||:::
    872 lyLeuLysLysArgThrLeuThrLeuGlyAsnIleSerVal......... 885

   1111 GCAAGAGAACCTGTTTATCATGCTGCAGGTGGTGTCAACAGTTATCGACC 1160
             |||                  :::|||:::      :::
    886 ......GluGlyAsnLeuSerLeuThrGlyAlaAsnAlaAsnIleValGl 900

   1161 CAGACTGAATAATGGAGAAAATATTTCCTTTATTGACGAAGGAAAAGGCG 1210
          :::|||:::   :::|||:::   :::|||      |||::::::   :
    900 yAsnLeuSerIleAlaGluAspSerThrPheLysGlyGluAlaSerAspA 917

   1211 AATTGATACTT........................ACCAGCAACATCAAT 1236
         ::|||   :::                         |||:::|||||||||
    917 snLeuAsnIleThrGlyThrPheThrAsnAsnGlyThrAlaAsnIleAsn 933

   1237 CAAGGTGCTGGAGGATTATATTTCCAAGGAGATTTTACGGTCTCGCCTGA 1286
                 |||   :::   :::||||||||||:::
    934 IleLysGlnGlyValValLysLeuGlnGlyAspIleIle........... 946

   1287 AAATAACGAAACTTGGCAAGGCGCGGGCGTTCATATCAGTGAAGACAGTA 1336
                        :::   :::|||::::::|||:::   :::::::
    947 ................AsnLysGlyGlyLeuAsnIleThrThrAsnAlaS 958

   1337 CCGTTACTTGGAAA......GTAAACGGC...GTGGCAAACGACCGCCTG 1377
        ::   |||   |||          :::||||||   ::::::|||:::
    958 erGlyThrGlnLysThrIleIleAsnGlyAsnIleThrAsnGlu...... 972

   1378 TCCAAAATCGGCAAAGGCACGCTG.........CACGTTCAAGCCAAAGG 1418
                      ||||||   |||          :::::::::|||   ::
    973 ............LysGlyAspLeuAsnIleLysAsnIleLysAlaAspAl 985

   1419 GGAAAACCAA......GGCTCGATCAGCGTGGGCGACGGTACAGTCATTT 1462
         :|||   |||        |||:::||||||      :::|||:::::   :
    985 aGluIleGlnIleGlyGlyAsnIleSerGlnLysGluGlyAsnLeuThrI 1002

   1463 TGGATCAGCAGGCAGACGATAAAGGCAAA......AAACAAGCCTTTAGT 1506
        ::          ::::::||||||   :::       :::|||
   1002 le.........SerSerAspLysValAsnIleThrAsnGlnIleThrIle 1015

   1507 GAAATCGGCTTGGTCAGCGGCAGGGGT....................... 1533
        :::   |||:::   :::|||||||:::
   1016 LysAlaGlyValGluGlyGlyArgSerAspSerSerGluAlaGluAsnAl 1032

   1534 .......ACGGTGCAACTGAATGCCGATAATCAGTTCAACCCCGACAAAC 1576
                |||:::|||   :::  ..:::   ::::::      |||
   1032 aAsnLeuThrIleGln...ThrLysGluLeuLysLeuAlaGlyAspLeuA 1048

   1577 TCTATTTCGGCTTTCGCGGCGGACGTTTGGATTTA...AACGGGCATTCG 1623
           |||||||:::   :::::::::           ||||||   :::
   1048 snIleSerGlyPheAsnLysAlaGluIleThrAlaLysAsnGlySerAsp 1064

   1624 CTTTCGTTCCACCGTATTCAAAATACCGATGAAGGGGCGATGATTGTCAA 1673
        |||:::::::   :::   ::::::   :::      |||      |||::
```

```
1065 LeuThrIleGlyAsnAlaSerGlyGlyAsnAlaAspAlaLysLysValTh 1081

1674 CCACAATCAAGACAAAGAATCCACCGTTACC..............ATTA 1708
      :  ::::::  |||:::|||  ::::::             :::|
1081 rPheAspLysValLysAspSerLysIleSerThrAspGlyHisAsnValT 1098

1709 CAGGCAATAAAGATATTGCTACAACC..................GGCAAT 1740
      ||  |||:::::::::  |||:::                  ||||||
1098 hrLeuAsnSerGluValLysThrSerAsnGlySerSerAsnAlaGlyAsn 1114

1741 AACAACAGCTTGGAT.........AGCAAAAAAGAAATTGCCTACAACGG 1781
      :::||||||          |||  |||:::::::::  |||
1115 AspAsnSerThrGlyLeuThrIleSerAlaLysAspValThrValAsn.. 1130

1782 TTGGTTTGGCGAGAAAGATACGACCAAAACGAACGGGCGGCTCAACCTTG 1831
                      :::  :::  ::::::  :::|||:::
1131 ................AsnIleValThrSerHisLysThrIleAsnIle. 1141

1832 TTTACCAGCCCGCCGCAGAAGACCGCACCCTGCTGCTTTCCGGCGGAACA 1881
                 :::|||     ::::::   :::  ||||||
1142 ...........SerAlaAlaAlaGlyAsnValThrThrLysGluGlyThr 1154

1882 AATTTAAACGGC...............AACATCACGCAAACAAACGGCAA 1916
      ::::::|||:::              ::::::|||    ||||||
1155 ThrIleAsnAlaThrThrGlySerValGluValThrAlaGlnAsnGlyTh 1171

1917 ACTGTTTTTCAGCGGCAGACCAACACCGCACGCCTACAATCATTTAAACG 1966
      :::     :::|||:::  |||
1171 rIle......LysGlyAsnIleThr......................... 1177

1967 ACCATTGGTCGCAAAAAGAGGGCATTCCTCGCGGGGAA......ATCGTG 2010
              ||||||:::     :::        |||    ::::::
1178 ........SerGlnAsnValThrValThrAlaThrGluAsnLeuValThr 1191

2011 TGGGACAACGACTGGATCAACCGCACATTTAAAGCGGAAAACTTCCAAAT 2060
        :::|||     ||||||  |||      :::  |||::::::
1192 ThrGluAsnAlaValIleAsnAlaThrSerGlyThrValAsnIleSerTh 1208

2061 TAAAGGCGGACAGGCGGTGGTTTCCCGCAATGTTGCCAAAGTGAAAGGCG 2110
        |||  |||:::       ::::::  ::::::  :::::::|||:
1208 rLysThrGlyAsp......IleLysGlyGlyIleGluSerThrSerGlyA 1223

2111 ATTGGCATTTGAGCAATCACGCCCAAGCAGTTTTTGGTGTCGCACCGCAT 2160
      ::  :::::::::      ::::::::::::       |||:::
1223 snValAsnIleThrAlaSerGlyAsnThrLeu...LysValSerAsnIle 1238

2161 CAAAGCCACACAATCTGTACACGTTCGGACTGGACGGGTCTGACAAATTG 2210
        ::::::  :::  :::  :::|||     :::||||||:::
1239 ThrGlyGlnAspValThrValThrAlaAspAlaGlyAlaLeuThrThrTh 1255

2211 TGTCGAAAAAAACCATTACCGACGATAAAGTGATTGCTTCATTG...ACTA 2257
        :::  :::||||||:::          |||::::::   ||||
1255 rAlaGlySerThrIleSerAlaThrThrGlyAsnAlaAsnIleThrThrL 1272

2258 AGACC...GACATCAGCGGCAATGTCGATCTTGCCGAT...CACGCTCAT 2301
      |||||  ||||||:::|||:::|||:::  ::::::      :::
1272 ysThrGlyAspIleAsnGlyLysValGluSerSerSerGlySerValThr 1288

2302 TTAAATCTCACAGGGCTTGCCACACTCAACGGCAATCTTAGTGCAAATGG 2351
      |||       ||||||  :::       ||||||:::|||:::|||
1289 LeuValAlaThrGlyAlaThrLeuAlaValGlyAsnIleSerGlyAsnTh 1305

2352 CGATACA..................CGTTATACAGTCAGCCACAACGCCA 2383
        |||                 :::  |||  :::  ::::::|
1305 rValThrIleThrAlaAspSerGlyLysLeuThrSerThrValGlySerT 1322

2384 CCCAAAACGGCAACCTTAGCCTCGTGGGCAATGCCCAAGCAACATTTAAT 2433
      ||  ||||||:::  |||::::::  ::::::|||:::
1322 hrIleAsnGlyThrAsnSerValThrThrSerSerGlnSerGlyAspIle 1338

2434 CAAGCCACATTAAACGGCAACACATCGGCTTCGGGCAATGCTTCATTTAA 2483
      ::::::|||::::::|||||||||                    ||
1339 GluGlyThrIleSerGlyAsnThr.....................ValAs 1348

2484 TCTAAGCGACCACGCCGTACAAAACGGCAGTCTGACGCTTTCCGGCAACG 2533
      |::::::            :::::::::|||:::||||||::::::::::::|
1348 nValThr.........AlaSerThrGlyAspLeuThrIleGlyAsnSerA 1362

2534 CTAAGGCAAACGTAAGCCATTCCGCACTCAACGGTAATGTCTCCCTAGCC 2583
      |||||::::::::::::::::::|||
1362 laLysValGluAlaLysAsnGlyAla........................ 1370

2584 GATAAGGCAGTATTCCATTTTGAAAGCAGCCGCTTTACCGGACAA..... 2628
           |||::::::       ||||||:::::::::|||    |||
1371 ......AlaThrLeuThrAlaGluSerGlyLysLeuThrThrGlnThrGl 1385

2629 ..........ATCAGCGGCGGCAAGGATACGGCATTACACTTAAAAGACA 2668
                 |||::::::  :::|||:::|||       |||||||
1385 ySerSerIleThrSerSerAsnGlyGlnThrThrLeuThrAlaLysAspS 1402

2669 GCGAATGGACGCTGCCGTCAGGCACGGAATTAGGCAATTTAAACCTTGAC 2718
      ||                  :::     ||||||:::|||
1402 er....................SerIleAlaGlyAsnIleAsnAla... 1410

2719 AACGCCACCATTACACTCAATTCCGCCTATCGCCACGATGCGGCAGGGGC 2768
         |||::::::|||||||||:::
1411 ...AlaAsnValThrLeuAsnThr.......................... 1417

2769 GCAAACCGGCAGTGCGACAGATGCGCCGCGCCGCCGTTCGCGCCGTTCGC 2818
         |||||||:::
1418 ....ThrGlyThr..................................... 1420

2819 GCCGTTCCCTATTATCCGTTACACCGCCAACTTCGGTAGAATCCCGTTTC 2868
                 |||::::::|||    :::::::::::::::
1421 ...........LeuThrThrThrGlyAspSerLysIleAsnAlaThrSer 1433

2869 AACACGCTGACGGTAAACGGCAAATTGAACGGTCAGGGAACATTCCGCTT 2918
      :::|||||||||:::|||:::|||
1434 GlyThrLeuThrIleAsnAlaLys.......................... 1441

2919 TATGTCGGAACTCTTCGGCTACCGCAGCGACAAATTGAAGCTGGCGGAAA 2968
                        :::  ||||||     |||    |
1442 ........................AspAlaLysLeuAspGlyAlaAlaS 1450

2969 GTTCCGAAGGCACTTACACCTTGGCGGTCAACAATACCGGCAACGAACCT 3018
      ||::::::  |||   :::  |||:::|||  :::|||:::
1450 erGlyAspArgThrValValAsnAlaThrAsnAlaSerGlySer...... 1464

3019 GCAAGCCTCGAACAATTGACGGTAGTGGAAGGAAAAGACAACAAACCGCT 3068
                               |||:::    :::
1465 ............................GlyAsnValThrAlaLysTh 1471

3069 GTCCGAAAACCTTAATTTCACCCTGCAAAACGAACACGTCGATGCCGGCG 3118
      |||:::::::::|||:::|||  :::  :::  :::::::::
1471 rSerSerSerValAsnIleThrGlyAspLeuAsnThrIleAsnGly.... 1486

3119 CGTGGCGTTACCAACTCATCCGCAAAGACGGCGAGTTCCGCCTGCATAAT 3168
               ::::::|||  ::::::|||              :::|||
1487 ........LeuAsnIleIleSerGluAsnGly............ArgAsn 1496

3169 CCGGTCAAAGAACAAGAGCTTTCCGACAAACTCGGCAAGGCAGAAGCCAA 3218
         |||                    :::|||        |||
1497 ThrVal.....................ArgLeuArgGlyLysGluIleAs 1506

3219 AAAACAGGCGGAAAAAGACAACGCGCAAAGCCTTGACGCGCTGATTGCGG 3268
         :::   :::  ::::::  |||::::::  :::|||   |
1506 pValLysTyrIleGlnProGlyValAlaSerValGluGluValIleGluA 1523

3269 CCGGGCGCGATGCCGTCGAAAAGACAGAAAGCGTTGCCGAACCGGCCCGG 3318
      ||   |||   ::::::|||||||:::::::::::::::::      |||
1523 laLysArg...ValLeuGluLysValLysAspLeuSerAspGluGluArg 1538

3319 CAGGCAGGCGGGGAAAATGTCGGCATTATGCAGGCGGAGGAAGAGAAAAA 3368
      ::::::  ::::::  :::|||:::     :::              ::
1539 GluThrLeuAlaLys...LeuGlyValSerAlaVal............Ar 1550
```

```
3369 ACGGGTGCAGGCGGATAAAGACACCGCCTTGGCGAAACAGCGCGAAGCGG 3418
       :  |||:::   ::::::      ::::::      |||:::|||
1550 gPheValGluProAsnAsnAlaIleThrValAsnThrGlnAsnGluPheT 1567

3419 AAACCCGGCCGGCTACC 3435
        |||:::|||::::::
1567 hrThrLysProSerSer 1572
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:AAR63506

```
seq_documentation_block:
ID   AAR63506 standard; Protein; 1477 AA.
XX
AC   AAR63506;
XX
DT   25-JUN-1995  (first entry)
XX
DE   Haemophilus high molecular weight protein HMW2.
XX
KW   High molecular weight protein; HMW2; protective vaccine; otitis;
KW   sinusitis; bronchitis; Hib; ss.
XX
OS   Haemophilus.
XX
PN   WO9421290-A.
XX
PD   29-SEP-1994.
XX
PF   15-MAR-1994;   94WO-US02550.
XX
PR   16-MAR-1993;   93US-0038682.
XX
PA   (BARE/) BARENKAMP S J.
PA   (SGEM/) ST GEME J W.
XX
PI   Barenkamp SJ,  St GEME JW;
XX
DR   WPI; 1994-316665/39.
DR   Q-PSDB; Q72294.
XX
PT   New immunogenic high mol. wt. proteins of non typeable
PT   Haemophilus - useful in protective vaccines
XX
PS    Claim 3; Page 36; 127pp; English.
XX
CC   The HMW2 protein encoded by this sequence is useful in a vaccine to
CC   protect against disease caused by non-typeable Haemophilus which are
CC   not controlled by H. influenzae type b (Hib) vaccines.  The encoded
CC   protein can also be used as a carrier for protective Hib
CC   polysaccharide (in a conjugate vaccin against meningitis) or for
CC   other antigens, haptens, etc.
XX
SQ   Sequence   1477 AA;

alignment_scores:
          Quality:  287.00          Length:   1437
            Ratio:  0.415             Gaps:     66
 Percent Similarity:  48.156   Percent Identity: 19.415

alignment_block:
 US-09-303-518D-649 x AAR63506   ..

 Align seg 1/1   to: AAR63506   from: 1   to: 1477

 163 CGCGACTTTGCCGAAAATAAAGGCAAGTTTGCAGTCGGGGCGAAAGATAT 212
      :::|||           |||   :::   |||:::::::::::   |||||
 158 LysAspAlaIleIleAsnThrAsnGlyPheThrAlaSerThrLeuAspIl 174

 213 TGAGGTTTACAACAAAAAAGGGGAGTTGGTCGGCAAATCAATGACAAAA. 261
      |:::        |||   |||:::::              :::::||||||
 174 eSerAsnGluAsnIleLysAlaArgAsnPheThrPheGluGlnThrLysA 191

 262 ...........GCCCCGATGATTGATTTTTCTGTG......GTGTCGCGT 294
                |||   :::::::::   ::::::      |||::::::
 191 spLysAlaLeuAlaGluIleValAsnHisGlyLeuIleThrValGlyLys 207

 295 AACGGCGTGGCGGCATTGGTGGGC...........GATCAATATATTGT 332
     :::|||   :::   |||:::|||            ::::::   :::::
 208 AspGlySerValAsnLeuIleGlyGlyLysValLysAsnGluGlyValIl 224

 333 GAGCGTGGCACATAACGGCGGCTATAACAACGTTGATTTTGGT....... 375
     :||||||        |||||||||      ::::::      |||
 224 eSerVal......AsnGlyGlySerIleSerLeuLeuAlaGlyGlnLysI 239

 376 ........GCGGAAGGAAGAAATCCCGATCAACATCGTTTTACTTATAAA 417
                 ::::::      ||||||          :::||||||:::
 239 leThrIleSerAspIleIleAsnPro.........ThrIleThrTyrSer 252

 418 ATTGTGAAACGGAATAATTATAAAGCAGGGACTAAAGGCCATCCTTAT.. 465
     |||:::      :::|||   :::|||   :::   |||      :::
 253 IleAlaAlaProGluAsn...GluAlaValAsnLeuGlyAspIlePheAl 268

 466 ....GGCGGCGATTATCATATGCCGCGTTTGCATAAATTTGTCACAGATG 511
         ||||||:::   :::
 268 aLysGlyGlyAsnIleAsn............................. 274

 512 CAGAACCTGTTGAAATGACCAGTTATATGGATGGGCGGAAATATATCGAT 561

 274 .................................................. 274

 562 CAAAATAATTACCCTGACCGTGTTCGTATTGGGGCAGGCAGGCAATATTG 611
                         ||||||   ::::::   |||:::
 275 ....................ValArgAlaAlaThrIleArgAsnGlnGl 284

 612 GCGATCTGATGAAGATGAGCCCAATAACCGCGAAAGTTCATATCATATTG 661
      :::   :::   |||:::   ::::::   :::|||:::      :::
 284 yLysLeuSerAlaAspSerValSerLysAspLysSerGlyAsnIleValL 301

 662 CAAGTGCGTATTCTTGG......CTCGTTGGTGGCAATACCTTTGCACAA 705
      ||||||   :::            :::||||||         ||||||
 301 euSerAlaLysGluGlyGluAlaGluIleGlyGlyValIleSerAlaGln 317

 706 AAT............................................... 708
     |||
 318 AsnGlnGlnAlaLysGlyGlyLysLeuMetIleThrGlyAspLysValTh 334

 709 ..............................GGATCAGGTGGTGGCACAG 727
                                   |||:::   ||||||   :
 334 rLeuLysThrGlyAlaValIleAspLeuSerGlyLysGluGlyGlyGluT 351

 728 TCAACTTAGGTAGTGAAAAAATTAAACATAGCCCATATGGTTTT...... 771
      ::   ||||||:::::::::      ::::::      |||:::
 351 hrTyrLeuGlyGlyAspGluArgGlyGluGlyLysAsnGlyIleGlnLeu 367

 772 ...............TTACCAACAGGAGGCTCATTTGGCGACAGTGGCTC 806
                    |||      |||:::::::::::   ||||||
 368 AlaLysLysThrSerLeuGluLysGlySerThrIleAsnValSerGly.. 383

 807 ACCAATGTTTATCTATGATGCCCAAAAGCAAAAG................ 840
                              |||:::|||
 384 ........................LysGluLysGlyGlyArgAlaIleV 392

 841 ..TGG............TTAATTAATGGGGTATTGCAAACG...GGCAAC 873
       |||             ||||||:::|||   :::::::::   |||:::
 392 alTrpGlyAspIleAlaLeuIleAspGlyAsnIleAsnAlaGlnGlySer 408

 874 CCCTATATAGGAAAAAGCAATGGCTTC....................... 900
           |||:::|||::::::||||||
 409 GlyAspIleAlaLysThrGlyGlyPheValGluThrSerGlyHisTyrLe 425

 901 ................CAGCTGGTTCGT...AAAGATTGGTTCTATGATG 931
                     :::|||:::   |||:::|||:::   |||
 425 uSerIleAspSerAsnAlaIleValLysThrLysGluTrpLeuLeuAspP 442

 932 AAATCTTTGCTGGAGATACCCATTCAGTATTCTACGAACCACGTCAAAAT 981
```

```
 442 ro...........AspAspValThrIleGluAlaGluAspProLeuArg 454

 982 GGGAAATACTCTTTTAACGACGAT...AATAATGGCACAGGAAAAATCAA 1028
 455 AsnAsnThrGlyIleAsnAspGluPheProThrGlyThrGlyGluAlaSe 471

1029 TGCCAAACATGAACACAATTCTCTGCCTAATAGATTAAAAACACGAACCG 1078
 471 rAspProLysLysAsnSerGluLeuLysThrThrLeuThrAsnThrThrI 488

1079 TTCAA......TTGTTTAATGTT...TCTTTATCCGAGACAGCAAGAGAA 1119
 488 leSerAsnTyrLeuLysAsnAlaTrpThrLeuAsnIleThrAlaSerArg 504

1120 CCTGTTTATCATGCTGCAGGTGGTGTCAACAGTTATCGACCCAGACTGAA 1169
 505 LysLeu..............ThrValAsnSer.........SerIleAs 513

1170 TAATGGAGAAAATATTTCCTTTATTGACGAAGGAAAAGGCGAATTGATAC 1219
 513 nIleGlySerAsnSerHisLeuIleLeuHisSerLysGly.......... 526

1220 TTACCAGCAACATCAATCAAGGTGCTGGAGGATTATATTTCCAAGGAGAT 1269
 527 ................GlnArgGlyGlyGlyValGlnIleAspGlyAsp 537

1270 TTTACGGTCTCGCCTGAAAATAACGAAACTTGGCAAGGCGCGGGCGTTCA 1319
 538 IleThrSerLysGlyGlyAsnLeuThrIleTyrSerGlyGlyTrpValAs 554

1320 TATCAGTGAAGACAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAAACG 1369
 554 pValHisLysAsnIleThrLeu........................... 561

1370 ACCGCCTGTCCAAAATCGGCAAAGGCACGCTGCACGTTCAAGCCAAAGGG 1419
 561 .................................................. 561

1420 GAAAACCAAGGCTCGATCAGCGTGGGCGACGGTACAGTCATTTTGGATCA 1469
 562 ...AspGlnGlyPheLeuAsnIleThrAlaAlaSerValAlaPhe...Gl 576

1470 GCAGGCAGACGATAAAGGCAAAAAACAAGCCTTTAGTGAAATCGGCTTGG 1519
 576 uGlyGlyAsnAsnLysAlaArgAspAlaAlaAsnAlaLysIle....... 590

1520 TCAGCGGCAGGGGTACGGTGCAACTGAATGCCGATAATCAGTTCAACCCC 1569
 591 ..ValAlaGlnGlyThrValThrIleThrGlyGluGlyLys......... 603

1570 GACAAACTCTATTTCGGCTTTCGCGGCGGACGTTTGGATTTAAACGGGCA 1619
 604 ..............AspPheArgAlaAsnAsnValSerLeuAsnGly.. 614

1620 TTCGCTTTCGTTCCACCGTATTCAAAATACCGATGAAGGGGCGATGATTG 1669
 615 ..........................ThrGlyLysGlyLeuAsnIleI 622

1670 TC...........AACCACAATCAAGACAAAGAATCCACCGTTACCATT 1707
 622 leSerSerValAsnAsnLeuThrHisAsnLeuSerGlyThrIleAsnIle 638

1708 ACAGGCAATAAAGATATTGCTACAACCGGCAATAACAACAGCTTGGATAG 1757
 639 SerGlyAsnIleThrIleAsnGlnThr....................Th 648

1758 CAAAAAAGAAATTGCCTACAACGGTTGGTTTGCCGAGAAAGATACG.... 1803
 648 rArgLysAsnThrSerTyr......TrpGlnThrSerHisAspSerHisT 663

1804 ...................ACCAAAACGAACGGGCGGCTCAACCTTGTT 1833

 663 rpAsnValSerAlaLeuAsnLeuGluThrGlyAlaAsnPheThrPheIle 679

1834 TACCAGCCCGCCGCAGAAGACCGCACCCTGCTGCTT...........TC 1871
 680 LysTyrIleSerSerAsnSerLysGlyLeuThrThrGlnTyrArgSerSe 696

1872 CGGCGGAACAAATTTAAACGGCAACATCACGCAAACAAACGGCAAACTGT 1921
 696 rAlaGlyValAsnPheAsnGly...........ValAsnGlyAsnMetS 709

1922 TTTTCAGCGGCAGACCAACACCGCACGCCTACAATCATTTA......AAC 1965
 709 erPheAsnLeuLysGluGlyAlaLysValAsnPheLysLeuLysProAsn 725

1966 GACCATTGGTCGCAAAAAGAGGGCATTCCT................... 1995
 726 GluAsnMetAsnThrSerLysProLeuProIleArgPheLeuAlaAsnIl 742

1996 ............CGCGGGGAAATCGTGTGGGACAACGACTGGATCAACC 2032
 742 eThrAlaThrGlyGlyGlySerValPhePheAsp.........IleTyrA 756

2033 GCACATTTAAAGCGGAAAACTTCCAAATTAAAGGCGGACAGGCGGTGGTT 2082
 756 laAsnHisSerGlyArgGlyAlaGluLeuLysMetSerGluIleAsnIle 772

2083 TCCCGCAATGTTGCCAAAGTGAAAGGCGATTGGCATTTGAGCAATCACGC 2132
 773 SerAsnGly..............AlaAsnPheThrLeuAsnSerHisVa 784

2133 CCAAGCAGTTTTTGGTGTCGCACCGCATCAAAGCCACACAATC...TGTA 2179
 784 lArgAlaAspAspAlaPheLysIleAsnLysAspLeuThrIleAsnAlaT 801

2180 CACGTTCGGACTGGACGGGTCTGACAAATTGTGTCGAAAAAACCATTACC 2229
 801 hrAsnSerAsnPhe.................................. 805

2230 GACGATAAAGTGATTGCTTCATTGACTAAGACCGACATCAGCGGCAATGT 2279
 806 .................SerLeuArgGlnThrLysAspAspPheTyrAs 816

2280 CGATCTTGCCGATCACGCTCATTTAAATCTCACAGGGCTTGCCACACTCA 2329
 816 pGlyTyrAlaArgAsnAlaIleAsnSerThrTyrAsnIleSerIleLeuG 833

2330 ACGGCAATCTTAGTGCAAATGGCGATACACGTTATACAGTCAGCCACAAC 2379
 833 lyGlyAsnValThrLeuGlyGly.......................... 840

2380 GCCACCCAAAACGGCAACCTTAGCCTCGTGGGCAATGCCCAAGCAACATT 2429
 841 ......GlnAsnSerSerSerSerIleThrGlyAsn......IleThrIl 853

2430 TAATCAAGCC.........ACATTAAACGGCAACACATCGGCTTCGGGCA 2470
 853 eGluLysAlaAlaAsnValThrLeuGluAlaAsnAsnAlaProAsnGlnG 870

2471 ATGCTTCATTTAATCTAAGCGACCACGCCGTACAAAACGGCAGTCTGACG 2520
 870 ln.........AsnIleArgAspArgValIleLysLeuGlySerLeuLeu 883

2521 CTTTCCGGCAACGCTAAGGCAAACGTAAGCCATTCCGCACTCAACGGTAA 2570
 884 ValAsnGlySerLeuSerLeuThrGlyGluAsnAlaAspIleLysGlyAs 900

2571 TGTCTCCCTAGCCGATAAGGCAGTATTCCAT.........TTTGAAAGCA 2611
 900 nLeuThrIleSerGluSerAlaThrPheLysGlyLysThrArgAspThrL 917

2612 GCCGCTTTACCGGACAAATCAGCGGCGGCAAGGATACGGCATTACACTTA 2661
 917 euAsnIleThrGlyAsnPheThrAsnAsnGlyThrAlaGluIleAsnIle 933
```

```
2662 AAAGACAGCGAATGGACGCTGCCGTCAGGCACGGAATTAGGCAATTTAAA 2711
         ::::::          |||   :::   |||:::   |||:::|||||
 934 ThrGlnGlyValValLysLeuGlyAsnValThrAsnAspGlyAspLeuAs 950

2712 CCTTGACAACGCCACCATTACACTCAATTCCGCCTATCGCCACGATGCGG 2761
      |:::   :::
 950 nIleThrThr........................................ 953

2762 CAGGGGCGCAAACCGGCAGTGCGACAGATGCGCCGCGCCGCCGTTCGCGC 2811
                                               :::::::::
 954 ..........................................HisAlaLys 956

2812 CGTTCGCGCCGTTCCCTATTATCCGT....TACACCGCCAACTTCGGTAG 2857
      |||:::::|||||||::::::::                 :::    |||
 957 ArgAsnGlnArgSerIleIleGlyGlyAspIleIleAsnLysLysGlySe 973

2858 AATCCCGTTTCAACACGCTGACGGTAAACGGCAAATTGAACGGTCAGGGA 2907
        :::   :::       :::::::::      :::|||:::   :::||||
 973 rLeuAsnIleThrAspSerAsnAsnAspAlaGluIleGlnIleGlyGlyA 990

2908 ACATTCCGCTTTATGTCGGAACTCTTCGGCTACCGCAGCGACAAATTGAA 2957
      |||||            |||:::|||    :::      |||||||||:::::
 990 snIleSerGlnLysGluGlyAsnLeuThrIle.SerSerAspLysIleAs 1006

2958 GCTGGCGGAAAGTTCCGAAGGCACTTACACCTTGGCGGTCAACAATACCG 3007
      :::::::
1006 nIleThr........................................... 1008

3008 GCAACGAACCTGCAAGCCTCGAACAATTGACGGTAGTGGAAGGCAAAAGAC 3057
                            :::|||:::|||:::   :::|||   |||
1009 ...................LysGlnIleThrIleLysLysGlyIleAsp 1018

3058 AACAAACCGCTGTCC............GAAAACCTTAATTTCACCCTGCA 3095
      ::::::        |||           :::|||    |||:::|||:::::
1019 GlyGluAspSerSerSerAspAlaThrSerAsnAlaAsnLeuThrIleLy 1035

3096 AAACGAACACGTCGAT....................GCCGGCGCGTGGC 3124
      ::::::::::::                           :::|||      :
1035 sThrLysGluLeuLysLeuThrGluAspLeuSerIleSerGlyPheAsnL 1052

3125 GTTACCAACTCATCCGCAAAGACGGC...GAGTTCCGCCTGCATAATCCG 3171
      ::   ::::::       |||||||||   ::::::   :::    |||
1052 ysAlaGluIleThrAlaLysAspGlyArgAspLeuThrIleGlyAsnSer 1068

3172 GTCAAAGAACAAGAGCTTTCCGACAAACTCGGCAAGGCAGAAGCCAAAAA 3221
                         :::|||       :::    |||||||||   ||
1069 ..................AsnAspGlyAsnSerGlyAlaGluAla...Ly 1078

3222 ACAGGCGGAAAAAGACAACGCGCAAAGCCTTGACGCGCTGATTGCGGCCG 3271
      |    :::         :::|||::::::        |||:::    |||:::|||
1078 sThrValThrPheAsnAsnValLys......AspSerLysIleSerAlaA 1093

3272 GGCGCGATGCCGTCGAAAAGACAGAAAGCGTTGCCGAACCGGCCCGGCAG 3321
                  |||          :::::::::|||          :::   :::
1093 spGlyHisAsnValThrLeuAsnSerLysValLysThrSerSerSerAsn 1109

3322 GCAGGCGGGGAAAATGTCGGCATTATGCAGGCGGAGGAAGAGAAAAAACG 3371
      :::|||   |||:::
1110 GlyGlyArgGluSer.................................. 1114

3372 GGTGCAGGCGGATAAAGACACCGCCTTGGCGAAACAGCGCGAAGCGGAAA 3421
           ::::::|||:::||||||:::|||
1115 ....AsnSerAspAsnAspThrGlyLeu..................... 1122

3422 CCCGGCCGGCTACCACCGCCTTCCCCCGCGCCCGCCGCGCCCGCCGGGAT 3471
                      |||          :::         ::::::::::::|||
1123 ..............ThrIleThrAlaLysAsnValGluValAsnLysAsp 1134

3472 TTGCCGCAACTGCAACCCCAACCGCAGCCCCAACCGCAGCGCGACCTGAT 3521
      :::    :::|||
1135 IleThrSerLeu...................................... 1138

3522 CAGCCGTTATGCCAATAGCGGTTTGAGTGAATTTTCCGCCACGCTCAACA 3571
         :::    :::|||            ||||||       :::|||  ::::::
1139 ....LysThrValAsnIleThrAlaSerGluLysValThrThr.ThrAla 1153

3572 GCGTTTTCGCCGTACAGGACGAATTAGACCGCGTATTTGCCGAAGACCGC 3621
      :::   :::           ||||||      |||    :::            ::
1154 GlySerThrIleAsnAlaThrAsnGlyLysAlaSerIleThrThrLysTh 1170

3622 CGCAACGCCGTTTGGACAAGCGGCATCCGGGACACCAAACACTACCGTTC 3671
      ::::         |||      :::::::::|||    :::::::::::
1170 rGlyAspIleSerGlyThrIleSerGlyAsnThrValSerValSerAlaT 1187

3672 GC...AAGATTTCCGCGCCTACCGCCAACAAACCGACCTGCGCCAAATCG 3718
             :::::::::   :::::::::|||      :::    ::::::
1187 hrGlyAspLeuThrThrLysSerGlySerLysIleGluAlaLysSerGly 1203

3719 GTATGCAGAAAAACCTCGGCAGCGGGCGCGTCGGCATCCTGTTTTCGCAC 3768
         ::::::    |||||||||:::|||:::    ::::::           ||
1204 GluAlaAsnValThrSerAlaThrGlyThrIleGlyGly.........Th 1217

3769 AACCGGACCGAAAACACCTTCGACGACGGCATCGGCA............. 3805
      |    :::    :::|||    ::::::|||||||:::|||
1217 rIleSerGlyAsnThrValAsnValThrAlaAsnAlaGlyAspLeuThrV 1234

3806 ......................ACTCGGCACGGCTTGCCCACGGCGCCG 3832
                               :::    :::|||      |||
1234 alGlyAsnGlyAlaGluIleAsnAlaThrGluGlyAlaAlaThrLeuThr 1250

3833 TTTTCGGGCAATACGGCATCGACA......GGTTCTACATCGGCATCAGC 3876
         :::|||||||||    :::|||     ||||||:::    :::|||::
1251 AlaThrGlyAsnThrLeuThrThrGluAlaGlySerSerIleThrSerTh 1267

3877 GCGGGCGCGGGTTTTAGCAGCGGCAGCCTTTCAGACGGCATCGGAGGCAA 3926
      ::::::::::|||            |||      |||::::::|||   |||
1267 rLysGlyGlnValAspLeuLeuAla......GlnAsnGlySerIleAlaG 1282

3927 AATCCGCCGCCGCGTGCTGCATTACGGCATTCAGGCACGATACCGCGCCG 3976
         |||      |||:::   :::|||                    |||
1282 lySerIleAsnAlaAlaAsnValThrLeuAsnThrThrGlyThrLeuThr 1298

3977 GTTTCGGCGGATTCGGCATCGAACCGCACATCGGCGCAACGCGCTATTTC 4026
      :::    |||   |||       :::    |||||||:::       :::|||::
1299 ThrValAlaGlySerAspIleLysAlaThrSerGlyThrLeuValIleAs 1315

4027 GTCCAAA 4033
      :::|||
1315 nAlaLys 1317
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:AAR41724

```
seq_documentation_block:
ID   AAR41724 standard; Protein; 1477 AA.
XX
AC   AAR41724;
XX
DT   26-APR-1994  (first entry)
XX
DE   High molecular weight protein 2 (HMW2).
XX
KW   HMW; high molecular weight protein; virus; vaccine; influenza;
KW   epitope; immunity; haemophilus influenzae.
XX
OS   Haemophilus influenzae.
XX
PN   WO9319090-A.
XX
PD   30-SEP-1993.
XX
PF   16-MAR-1993;   93WO-US02166.
XX
PR   16-MAR-1992;   92GB-0005704.
XX
PA   (BARE/) BARENKAMP S J.
```

```
PA   (INRM ) INSERM INST NAT SANTE & RECH  MEDICALE.
XX
PI   Barenkamp SJ;
XX
DR   WPI; 1993-320683/40.
DR   N-PSDB; AAQ49507.
XX
PT   High molecular weight surface proteins - of non-typeable
PT   haemophilus which exhibit immunogenic properties
XX
PS   Claim 4; Figure 4; 100pp; English.
XX
CC   The isolation and purification of the high molecular weight protein
CC   enables the identification of the major protective epitopes of the
CC   protein by conventional epitope mapping.  These epitopes can then be
CC   synthesised using standard techniques and incorporated into fully
CC   synthetic or recombinant vaccines.
XX
SQ   Sequence   1477 AA;


alignment_scores:
          Quality:  285.50          Length:   1388
            Ratio:  0.432             Gaps:     61
 Percent Similarity:  47.622   Percent Identity: 19.380

alignment_block:
 US-09-303-518D-649 x AAR41724   ..

 Align seg 1/1  to: AAR41724  from: 1  to: 1477

    175 GAAAATAAAGGCAAGTTTGCAGTCGGGGCGAAAGATATTGAGGTTTACAA 224
         :::|||:::|||  ::::::|||::::::  :::|||::::::
    219 LysAsnGluGlyValIleSerValAsnGlyGlySerIleSerLeu..... 233

    225 CAAAAAAGGGGAGTTGGTCGGCAAATCAATGACAAAAGCC.......... 264
                     |||:::|||:::::::::|||   :::
    234 .............LeuAlaGlyGlnLysIleThrIleSerAspIleIleA 246

    265 ..CCGATGATTGATTTTTCTGTGGTGTCGCGTAACGGCGTGGCGGCATTG 312
          |||   |||   :::|||:::::::::  ::::::  |||:::
    246 snProThrIleThrTyrSerIleAlaAlaProGluAsnGluAlaValAsn 262

    313 GTGGGCGATCAATATATTGTGAGCGTGGCACATAACGGCGGCTATAACAA 362
         :::||||||              :::    :::||||||    ||
    263 LeuGlyAsp...............IlePheAlaLysGlyGlyAsnIleAs 274

    363 CGTT...........................GATTTTGGTGCGGAAGGAA 385
         ||||                            ::::::|||::::::
    274 nValArgAlaAlaThrIleArgAsnGlnGlyLysLeuSerAlaAspSerV 291

    386 GAAATCCCGATCAACATCGTTTTACTTATAAAATTGTGAAACGGAATAAT 435
          :::   |||:::                     :::|||:::::::::
    291 alSerLysAspLysSerGlyAsnIleValLeuSerAlaLysGluGlyGlu 307

    436 TATAAAGCAGGG...............ACTAAAGGCCATCCTTATGGCGG 470
          :::   |||                  :::   :::      |||||
    308 AlaGluIleGlyGlyValIleSerAlaGlnAsnGlnGlnAlaLysGlyGl 324

    471 CGATTATCATATGCCGCGTTTGCATAAATTTGTCACAGATGCAGAACCTG 520
         |     :::                  :::    :::       :
    324 yLysLeuMetIleThrGlyAspLysValThrLeuLysThrGlyAlaValI 341

    521 TTGAAATGACCAGTTATATGGATGGGCGGAAATATATCGATCAAAATAAT 570
         :::::::::::::            |||:::   |||:::     ::::::
    341 leAspLeuSerGlyLysGluGlyGlyGluThrTyrLeuGlyGlyAspGlu 357

    571 TACCCTGACCGTGTTCGTATTGGGGCAGGCAGGCAATATTGGCGATCTGA 620
                                 |||   |||::::::      :::
    358 Arg..................GlyGluGlyLysAsnGlyIleGlnLeuAl 368

    621 TGAAGATGAGCCCAATAACCGCGAAAGTTCATATCATATTGCAAGT.... 666
         :::             ::::::   |||:::   ::::::::::::

    368 aLysLysThrSerLeuGluLysGlySerThrIleAsnValSerGlyLysG 385

    667 ..............GCGTATTCTTGG...........CTCGTTGGTGGC 690
                      |||     |||             |||:::   |||
    385 luLysGlyGlyArgAlaIleValTrpGlyAspIleAlaLeuIleAspGly 401

    691 AATACCTTTGCACAAAATGGATCAGGTGGTGGCACAGTCAACTTAGGTAG 740
         |||        ||||||   |||||||||
    402 AsnIleAsnAlaGln...GlySerGly....................... 409

    741 TGAAAAAATTAAACATAGCCCATATGGTTTTTACCAACAGGAGGCTCAT 790
                                     :::    |||||||||
    410 .........................AspIleAlaLysThrGlyGlyPheV 418

    791 TTGGCGACAGTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCAAAAG 840
                 ||||||      :::||||||   |||
    418 alGluThrSerGlyHisAspLeuPheIleLysAsp............... 429

    841 TGGTTAATTAATGGGGTATTGCAAACGGGCAACCCCTATATAGGAAAAAG 890
                 |||:::::::::::::::
    430 .........AsnAlaIleValAspAla....................... 435

    891 CAATGGCTTCCAGCTGGTTCGTAAAGATTGGTTCTATGATGAAATCTTTG 940
                                   |||:::|||:::  |||
    436 ......................LysGluTrpLeuLeuAsp.......... 441

    941 CTGGAGATACCCATTCAGTATTCTACGAACCACGTCAAAATGGGAAATAC 990
           |||:::   |||:::      |||           :::::::::
    442 ..PheAspAsnValSerIleAsnAlaGluAspProLeuArgAsnAsnThr 457

    991 TCTTTTAACGACGAT...AATAATGGCACAGGAAAAATCAATGCCAAACA 1037
         ::::::||||||:::      :::||||||||||:::    :::
    458 GlyIleAsnAspGluPheProThrGlyThrGlyGluAlaSerAspProLy 474

   1038 TGAACACAATTCTCTGCCTAATAGATTAAAAACACGAACCGTTCAATTGT 1087
          ::::::::::::|||   :::   |||   :::   |||::::::    :
    474 sLysAsnSerGluLeuLysThrThrLeuThrAsnThrThrIleSerAsnT 491

   1088 TT.....................AATGTTTCTTTATCCGAGACAGCAAGA 1116
         ::                       |||::::::   |||:::
    491 yrLeuLysAsnAlaTrpThrMetAsnIleThrAlaSerArgLysLeuThr 507

   1117 GAACCTGTTTATCATGCTGCAGGTGGTGTCAACAGTTATCGACCCAGACT 1166
                                    |||||||||          ::
    508 ..........................ValAsnSer.........SerIl 512

   1167 GAATAATGGAGAAAATATTTCCTTTATTGACGAAGGAAAAGGCGAATTGA 1216
         :|||   |||:::|||      :::|||    ::::::|||||
    512 eAsnIleGlySerAsnSerHisLeuIleLeuHisSerLysGly....... 526

   1217 TACTTACCAGCAACATCAATCAAGGTGCTGGAGGATTATATTTCCAAGGA 1266
                                |||   :::|||||||:::    ::::::|||
    527 ....................GlnArgGlyGlyGlyValGlnIleAspGly 536

   1267 GATTTACGGTCTCGCCTGAAAATAACGAAACTTGGCAAGGCGCGGGCGT 1316
         |||:::|||   :::        |||         ::::::|||:::    ||
    537 AspIleThrSerLysGlyGlyAsnLeuThrIleTyrSerGlyGlyTrpVa 553

   1317 TCATATCAGTGAAGACAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAA 1366
         |   :::   ::::::   |||:::
    553 lAspValHisLysAsnIleThrLeu......................... 561

   1367 ACGACCGCCTGTCCAAAATCGGCAAAGGCACGCTGCACGTTCAAGCCAAA 1416

    561 .................................................. 561

   1417 GGGGAAAACCAAGGCTCGATCAGCGTGGGCGACGGTACAGTCATTTTGGA 1466
               :::||||||    :::::::::      :::::::|||    :::
    562 ......AspGlnGlyPheLeuAsnIleThrAlaAlaSerValAlaPhe.. 575

   1467 TCAGCAGGCAGACGATAAAGGCAAAAAACAAGCCTTTAGTGAAATCGGCT 1516
          :::   :::::::::|||::::::         |||    ::::::|||
    576 .GluGlyGlyAsnAsnLysAlaArgAspAlaAlaAsnAlaLysIle.... 590
```

```
1517 TGGTCAGCGGCAGGGGTACGGTGCAACTGAATGCCGATAATCAGTTCAAC 1566
             ::::::|||||||||   ::::::::::::::::::
 591 .....ValAlaGlnGlyThrValThrIleThrGlyGluGlyLys..... 603

1567 CCCGACAAACTCTATTTCGGCTTTCGCGGCGGACGTTTGGATTTAAACGG 1616
                          ||||||:::::::::::::::::||||||||
 604 ..................AspPheArgAlaAsnAsnValSerLeuAsnGl 614

1617 GCATTCGCTTTCGTTCCACCGTATTCAAAATACCGATGAAGGGGCGATGA 1666
     |                                       |||   :::|||      |
 614 y............................ThrGlyLysGlyLeuAsnI 621

1667 TTGTC............AACCACAATCAAGACAAAGAATCCACCGTTACC 1704
     ||:::               |||    :::::::::   ::::::|||::::::
 621 leIleSerSerValAsnAsnLeuThrHisAsnLeuSerGlyThrIleAsn 637

1705 ATTACAGGCAATAAAGATATTGCTACAACCGGCAATAACAACAGCTTGGA 1754
     |||::::||||||        |||        |||
 638 IleSerGlyAsnIleThrIleAsnGlnThr................... 647

1755 TAGCAAAAAAGAAATTGCCTACAACGGTTGGTTTGGCGAGAAAGATACG. 1803
      :::::::|||:::   ::::|||        |||     :::    |||:::
 648 .ThrArgLysAsnThrSerTyr......TrpGlnThrSerHisAspSerH 662

1804 ......................ACCAAAACGAACGGGCGGCTCAACCTT 1830
                              :::|||::::::::::::::::::
 662 isTrpAsnValSerAlaLeuAsnLeuGluThrGlyAlaAsnPheThrPhe 678

1831 GTTTACCAGCCCGCCGCAGAAGACCGCACCCTGCTGCTT........... 1869
      :::              ::::::::::::::    |||
 679 IleLysTyrIleSerSerAsnSerLysGlyLeuThrThrGlnTyrArgSe 695

1870 .TCCGGCGGAACAAATTTAAACGGCAACATCACGCAAACAAACGGCAAAC 1918
      |||::::|||::::|||::::||||||                  :::||||||||::::
 695 rSerAlaGlyValAsnPheAsnGly............ValAsnGlyAsnM 708

1919 TGTTTTTCAGCGGCAGACCAACACCGCACGCCTACAATCATTTAAACGAC 1968
      ::    |||
 708 etSerPhe......................................... 710

1969 CATTGGTCGCAAAAAGAGGGCATTCCTCGCGGGGAAATCGTGTGGGACAA 2018
           :::    |||||||||            :::        :::      ||
 711 ......AsnLeuLysGluGlyAlaLysValAsnPheLysLeuLysProAs 725

2019 CGACTGGATCAACCGCACA.................TTTAAAGCGGAAA 2050
      |:::    :::|||    :::                    |||    |||    |
 725 nGluAsnMetAsnThrSerLysProLeuProIleArgPheLeuAla...A 741

2051 ACTTCCAAATTAAAGGCGGACAGGCGGTGGTT.................. 2082
      ||:::             ||||||    :::|||
 741 snIleThrAlaThrGlyGlyGlySerValPhePheAspIleTyrAlaAsn 757

2083 ...TCCCGCAATGTTGCCAAAGTGAAA...GGCGATTGGCATTTGAGCAA 2126
         |||    :::      |||::::::|||      ::::::    ::::::|||||
 758 HisSerGlyArgGlyAlaGluLeuLysMetSerGluIleAsnIleSerAs 774

2127 TCACGCCCAAGCAGTTTTTGGTGTCGCACCGCATCAAAGCCACACAATCT 2176
      |    |||:::    ::::::                 :::||||||
 774 nGlyAlaAsnPheThrLeu..............AsnSerHis....... 783

2177 GTACACGTTCGGACTGGACGGGTCTGACAAATTGTGTCGAAAAAACCATT 2226
       :::|||:::|||     :::             |||       :::    ||||||
 784 ..ValArgGlyAspAspAlaPheLysIleAsn...LysAspLeuThrIle 798

2227 ACCGACGATAAAGTGATTGCTTCATTGACTAAGACCGACATCAGCGGCAA 2276
      :::        :::           ||||||     :::|||         :::
 799 AsnAlaThrAsnSerAsnPheSerLeuArgGlnThrLysAspAspPheTy 815

2277 TGTCGATCTTGCCGATCACGCTCATTTAAATCTCACAGGGCTTGCCACAC 2326
              |||   :::|||        :::        :::::::::    |
 815 rAspGlyTyrAlaArgAsnAlaIleAsnSerThrTyrAsnIleSerIleL 832

2327 TCAACGGCAATCTTAGTGCAAATGGCGATACACGTTATACAGTCAGCCAC 2376
     ||:::|||||||:::::     :::|||
 832 euGlyGlyAsnValThrLeuGlyGly....................... 840

2377 AACGCCACCCAAAACGGCAACCTTAGCCTCGTGGGCAATGCCCAAGCAAC 2426
                 ||||||:::::    |||::::::|||||||            ||
 841 .........GlnAsnSerSerSerIleThrGlyAsn......IleTh 852

2427 ATTTAATCAAGCC.........ACATTAAACGGCAACACATCGGCTTCGG 2467
     |::::::::::|||               ||||||::::::|||::::::     :::
 852 rIleGluLysAlaAlaAsnValThrLeuGluAlaAsnAsnAlaProAsnG 869

2468 GCAATGCTTCATTTAATCTAAGCGACCACGCCGTACAAAACGGCAGTCTG 2517
      :::                |||:::      |||:::::::::::::      |||||||||
 869 lnGln.........AsnIleArgAspArgValIleLysLeuGlySerLeu 882

2518 ACGCTTTCCGGCAACGCTAAGGCAAACGTAAGCCATTCCGCACTCAACGG 2567
       ::::::|||:::    :::      :::      :::::::::     ::::::||
 883 LeuValAsnGlySerLeuSerLeuThrGlyGluAsnAlaAspIleLysGl 899

2568 TAATGTCTCCCTAGCCGATAAGGCAGTATTCCAT.........TTTGAAA 2608
     ||||:::::::::::::::::|||:::|||                          ::::
 899 yAsnLeuThrIleSerGluSerAlaThrPheLysGlyLysThrArgAspT 916

2609 GCAGCCGCTTTACCGGACAAATCAGCGGCGGCAAGGATACGGCATTACAC 2658
      ::        ::::::||||||::::::::::::::          :::    ::::::
 916 hrLeuAsnIleThrGlyAsnPheThrAsnAsnGlyThrAlaGluIleAsn 932

2659 TTAAAAGACAGCGAATGGACGCTGCCGTCAGGCACGGAATTAGGCAATTT 2708
      :::    ::::::          |||    :::    |||:::       |||:::||
 933 IleThrGlnGlyValValLysLeuGlyAsnValThrAsnAspGlyAspLe 949

2709 AAACCTTGACAACGCCACCATTACACTCAATTCCGCCTATCGCCACGATG 2758
      ||||:::    :::.
 949 uAsnIleThrThr..................................... 953

2759 CGGCAGGGGCGCAAACCGGCAGTGCGACAGATGCGCCGCGCCGCCGTTCG 2808
                                                       ::::::
 954 ..........................................HisAla 955

2809 CGCCGTTCGCGCCGTTCCCTATTATCCGT....TACACCGCCAACTTCGG 2854
      :::|||::::::|||||||||:::::::::               :::    ||
 956 LysArgAsnGlnArgSerIleIleGlyGlyAspIleIleAsnLysLysGl 972

2855 TAGAATCCCGTTTCAACACGCTGACGGTAAACGGCAAATTGAACGGTCAG 2904
      |     :::     :::       :::::::::        :::|||:::      :::|
 972 ySerLeuAsnIleThrAspSerAsnAsnAspAlaGluIleGlnIleGlyG 989

2905 GGAACATTCCGCTTTATGTCGGAACTCTTCGGCTACCGCAGCGACAAATT 2954
      ||||||||                     |||:::|||     :::       |||||||||::
 989 lyAsnIleSerGlnLysGluGlyAsnLeuThrIle.SerSerAspLysIl 1005

2955 GAAGCTGGCGGAAAGTTCCGAAGGCACTTACACCTTGGCGGTCAACAATA 3004
      :::::::::
1005 eAsnIleThr....................................... 1008

3005 CCGGCAACGAACCTGCAAGCCTCGAACAATTGACGGTAGTGGAAGGAAAA 3054
                             :::|||:::|||:::     :::|||
1009 .......................LysGlnIleThrIleLysLysGlyIle 1017

3055 GACAACAAACCGCTGTCC............GAAAACCTTAATTTCACCCT 3092
     |||:::::        |||                 :::|||    |||:::|||::
1018 AspGlyGluAspSerSerSerAspAlaThrSerAsnAlaAsnLeuThrIl 1034

3093 GCAAAACGAACACGTCGAT....................GCCGGCGCGT 3121
      :::::::::::::::                                :::|||
1034 eLysThrLysGluLeuLysLeuThrGluAspLeuSerIleSerGlyPheA 1051

3122 GGCGTTACCAACTCATCCGCAAAGACGGC...GAGTTCCGCCTGCATAAT 3168
       :::    ::::::            |||||||||     ::::::     :::    |||
1051 snLysAlaGluIleThrAlaLysAspGlyArgAspLeuThrIleGlyAsn 1067

3169 CCGGTCAAAGAACAAGAGCTTTCCGACAAACTCGGCAAGGCAGAAGCCAA 3218
```

```
                           :::|||      :::   |||||||||
1068 Ser.................AsnAspGlyAsnSerGlyAlaGluAla.. 1077

3219 AAAACAGGCGGAAAAAGACAACGCGCAAAGCCTTGACGCGCTGATTGCGG 3268
      |||   :::      :::|||:::::      |||:::   |||:::|
1078 .LysThrValThrPheAsnAsnValLys......AspSerLysIleSerA 1092

3269 CCGGGCGCGATGCCGTCGAAAAGACAGAAAGCGTTGCCGAACCGGCCCGG 3318
      ||          |||      :::::::::|||         :::
1092 laAspGlyHisAsnValThrLeuAsnSerLysValLysThrSerSerSer 1108

3319 CAGGCAGGCGGGGAAAATGTCGGCATTATGCAGGCGGAGGAAGAGAAAAA 3368
      ::::::|||   |||:::
1109 AsnGlyGlyArgGluSer.............................. 1114

3369 ACGGGTGCAGGCGGATAAAGACACCGCCTTGGCGAAACAGCGCGAAGCGG 3418
             ::::::|||:::|||||||:::|||
1115 .......AsnSerAspAsnAspThrGlyLeu.................. 1122

3419 AAACCCGGCCGGCTACCACCGCCTTCCCCCGCGCCCGCCGCGCCCGCCGG 3468
                       |||         :::      ::::::::::::
1123 .................ThrIleThrAlaLysAsnValGluValAsnLys 1133

3469 GATTTGCCGCAACTGCAACCCCAACCGCAGCCCCAACCGCAGCGCGACCT 3518
      |||:::   :::|||
1134 AspIleThrSerLeu................................. 1138

3519 GATCAGCCGTTATGCCAATAGCGGTTTGAGTGAATTTTCCGCCACGCTCA 3568
             :::   :::|||           ||||||      :::||| :::
1139 .......LysThrValAsnIleThrAlaSerGluLysValThrThr.Thr 1152

3569 ACAGCGTTTTCGCCGTACAGGACGAATTAGACCGCGTATTTGCCGAAGAC 3618
      ::::::   :::           ||||||      |||   :::
1153 AlaGlySerThrIleAsnAlaThrAsnGlyLysAlaSerIleThrThrLy 1169

3619 CGCCGCAACGCCGTTTGGACAAGCGGCATCCGGGACACCAAACACTACCG 3668
       ::::::          |||      :::::::::|||   ::::::::::
1169 sThrGlyAspIleSerGlyThrIleSerGlyAsnThrValSerValSerA 1186

3669 TTCGC...AAGATTTCCGCGCCTACCGCCAACAAACCGACCTGCGCCAAA 3715
      ::         :::::::::   :::::::::|||      :::   :::
1186 laThrValAspLeuThrThrLysSerGlySerLysIleGluAlaLysSer 1202

3716 TCGGTATGCAGAAAAACCTCGGCAGCGGGCGCGTCGGCATCCTGTTTTCG 3765
      :::   ::::::   |||||||||:::|||:::   ::::::
1203 GlyGluAlaAsnValThrSerAlaThrGlyThrIleGlyGly........ 1216

3766 CACAACCGGACCGAAAACACCTTCGACGACGGCATCGGCA.......... 3805
       |||   :::   :::|||   ::::::||||||:::|||
1217 .ThrIleSerGlyAsnThrValAsnValThrAlaAsnAlaGlyAspLeuT 1233

3806 .........................ACTCGGCACGGCTTGCCCACGGCG 3829
                              :::   :::|||      |||
1233 hrValGlyAsnGlyAlaGluIleAsnAlaThrGluGlyAlaAlaThrLeu 1249

3830 CCGTTTTCGGGCAATACGGCATCGACA......GGTTCTACATCGGCATC 3873
          :::|||||||||   :::|||         ||||||:::   :::||
1250 ThrAlaThrGlyAsnThrLeuThrThrGluAlaGlySerSerIleThrSe 1266

3874 AGCGCGGGCGCGGGTTTTAGCAGCGGCAGCCTTTCAGACGGCATCGGAGG 3923
      |::::::::::::|||            |||      |||:::::::|||    |
1266 rThrLysGlyGlnValAspLeuLeuAla......GlnAsnGlySerIleA 1281

3924 CAAAATCCGCCGCCGCGTGCTGCATTACGGCATTCAGGCACGATACCGCG 3973
      ||    |||        |||:::   :::|||               |||
1281 laGlySerIleAsnAlaAlaAsnValThrLeuAsnThrThrGlyThrLeu 1297

3974 CCGGTTTCGGCGGATTCGGCATCGAACCGCACATCGGCGCAACGCGCTAT 4023
         :::   |||   |||         :::   ||||||:::      :::||
1298 ThrThrValAlaGlySerAspIleLysAlaThrSerGlyThrLeuValIl 1314

4024 TTCGTCCAAA 4033
      |::::::|||
```

```
1314 eAsnAlaLys 1317
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:AAR41728

```
seq_documentation_block:
ID   AAR41728 standard; Protein; 1477 AA.
XX
AC   AAR41728;
XX
DT   26-APR-1994  (first entry)
XX
DE   High molecular weight protein 2 (HMW2).
XX
KW   HMW; high molecular weight protein; virus; vaccine; influenza;
KW   epitope; immunity; haemophilus influenzae; gene cluster.
XX
OS   Haemophilus influenzae.
XX
PN   WO9319090-A.
XX
PD   30-SEP-1993.
XX
PF   16-MAR-1993;   93WO-US02166.
XX
PR   16-MAR-1992;   92GB-0005704.
XX
PA   (BARE/) BARENKAMP S J.
PA   (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
XX
PI   Barenkamp SJ;
XX
DR   WPI; 1993-320683/40.
DR   N-PSDB; AAQ49509.
XX
PT   High molecular weight surface proteins - of non-typeable
PT   haemophilus which exhibit immunogenic properties
XX
PS   Claim 4; Figure 4; 100pp; English.
XX
CC   The isolation and purification of the high molecular weight protein
CC   enables the identification of the major protective epitopes of the
CC   protein by conventional epitope mapping.  These epitopes can then be
CC   synthesised using standard techniques and incorporated into fully
CC   synthetic or recombinant vaccines.  This sequence is claimed to be
CC   the same as that given in AAR41724 (High molecular weight protein 2)
CC   although it does differ slightly.
XX
SQ   Sequence   1477 AA;

alignment_scores:
            Quality:  285.00          Length:   1441
              Ratio:  0.417             Gaps:     67
 Percent Similarity:  47.398   Percent Identity: 19.778

alignment_block:
 US-09-303-518D-649 x AAR41728   ..

 Align seg 1/1  to: AAR41728  from: 1  to: 1477

 163 CGCGACTTTGCCGAAAATAAAGGCAAGTTTGCAGTCGGGGCGAAAGATAT 212
      :::|||          |||   :::   |||:::::::::::::   |||||
 158 LysAspAlaIleIleAsnThrAsnGlyPheThrAlaSerThrLeuAspIl 174

 213 TGAGGTTTACAACAAAAAAGGGGAGTTGGTCGGCAAATCAATGACAAAA. 261
      |:::      |||   |||::::::               ::::::||||||
 174 eSerAsnGluAsnIleLysAlaArgAsnPheThrPheGluGlnThrLysA 191

 262 ..........GCCCCGATGATTGATTTTTCTGTG......GTGTCGCGT 294
                 |||   :::::::::   ::::::         |||::::::
 191 spLysAlaLeuAlaGluIleValAsnHisGlyLeuIleThrValGlyLys 207

 295 AACGGCGTGGCGGCATTGGTGGGC...........GATCAATATATTGT 332
```

```
         :::|||   :::   |||:::|||              ::::::   ::::
 208 AspGlySerValAsnLeuIleGlyGlyLysValLysAsnGluGlyValIl 224

 333 GAGCGTGGCACATAACGGCGGCTATAACAACGTTGATTTTGGT....... 375
     :||||||        |||||||||       ::::::      |||
 224 eSerVal......AsnGlyGlySerIleSerLeuLeuAlaGlyGlnLysI 239

 376 ........GCGGAAGGAAGAAATCCCGATCAACATCGTTTTACTTATAAA 417
           ::::::    ||||||           :::|||||||:::
 239 leThrIleSerAspIleIleAsnPro.........ThrIleThrTyrSer 252

 418 ATTGTGAAACGGAATAATTATAAAGCAGGGACTAAAGGCCATCCTTAT.. 465
     |||:::        :::|||   :::|||   :::    |||    . :::
 253 IleAlaAlaProGluAsn...GluAlaValAsnLeuGlyAspIlePheAl 268

 466 ....GGCGGCGATTATCATATGCCGCGTTTGCATAAATTTGTCACAGATG 511
         |||||||:::    :::
 268 aLysGlyGlyAsnIleAsn............................... 274

 512 CAGAACCTGTTGAAATGACCAGTTATATGGATGGGCGGAAATATATCGAT 561

 274 .................................................. 274

 562 CAAAATAATTACCCTGACCGTGTTCGTATTGGGGCAGGCAGGCAATATTG 611
                           ||||||   ::::::    |||:::
 275 ....................ValArgAlaAlaThrIleArgAsnGlnGl 284

 612 GCGATCTGATGAAGATGAGCCCAATAACCGCGAAAGTTCATATCATATTG 661
       :::    :::    |||:::   ::::::    :::|||:::     :::
 284 yLysLeuSerAlaAspSerValSerLysAspLysSerGlyAsnIleValL 301

 662 CAAGTGCGTATTCTTGG......CTCGTTGGTGGCAATACCTTTGCACAA 705
       ||||||   :::                :::||||||       ||||||
 301 euSerAlaLysGluGlyGluAlaGluIleGlyGlyValIleSerAlaGln 317

 706 AAT............................................... 708
      |||
 318 AsnGlnGlnAlaLysGlyGlyLysLeuMetIleThrGlyAspLysValTh 334

 709 ..............................GGATCAGGTGGTGGCACAG 727
                                  |||:::   ||||||   :
 334 rLeuLysThrGlyAlaValIleAspLeuSerGlyLysGluGlyGlyGluT 351
                                                   |||
 728 TCAACTTAGGTAGTGAAAAAATTAAACATAGCCCATATGGTTTT...... 771
      ::    |||||||:::::::::       ::::::      |||:::
 351 hrTyrLeuGlyGlyAspGluArgGlyGluGlyLysAsnGlyIleGlnLeu 367

 772 ...............TTACCAACAGGAGGCTCATTTGGCGACAGTGGCTC 806
                  |||       |||::::::::::::    ||||||
 368 AlaLysLysThrSerLeuGluLysGlySerThrIleAsnValSerGly.. 383

 807 ACCAATGTTTATCTATGATGCCCAAAAGCAAAAG................ 840
                           |||:::|||
 384 ........................LysGluLysGlyGlyArgAlaIleV 392

 841 ..TGG............TTAATTAATGGGGTATTGCAAACG...GGCAAC 873
        |||             |||||||:::|||    :::::::::    |||:::
 392 alTrpGlyAspIleAlaLeuIleAspGlyAsnIleAsnAlaGlnGlySer 408

 874 CCCTATATAGGAAAAAGCAATGGCTTC....................... 900
          |||:::|||:::::|||||||
 409 GlyAspIleAlaLysThrGlyGlyPheValGluThrSerGlyHisTyrLe 425

 901 ................CAGCTGGTTCGT...AAAGATTGGTTCTATGATG 931
                    :::|||:::     |||:::|||:::    |||
 425 uSerIleAspSerAsnAlaIleValLysThrLysGluTrpLeuLeuAspP 442

 932 AAATCTTTGCTGGAGATACCCATTCAGTATTCTACGAACCACGTCAAAAT 981
                   |||      ::::::     |||          :::
 442 ro...........AspAspValThrIleGluAlaGluAspProLeuArg 454

 982 GGGAAATACTCTTTTAACGACGAT...AATAATGGCACAGGAAAAAATCAA 1028
     ::::::   ::::::|||||||:::     :::||||||||||:::     ::

 455 AsnAsnThrGlyIleAsnAspGluPheProThrGlyThrGlyGluAlaSe 471

1029 TGCCAAACATGAACACAATTCTCTGCCTAATAGATTAAAAACACGAACCG 1078
     :             ::::::::::::|||    :::    |||    :::    |||:
 471 rAspProLysLysAsnSerGluLeuLysThrThrLeuThrAsnThrThrI 488

1079 TTCAATTGTTT.....................AATGTTTCTTTATCCGAG 1107
     :::::   :::                     |||:::::::    |||:::
 488 leSerAsnTyrLeuLysAsnAlaTrpThrMetAsnIleThrAlaSerArg 504

1108 ACAGCAAGAGAACCTGTTTATCATGCTGCAGGTGGTGTCAACAGTTATCG 1157
                                     |||||||||
 505 LysLeuThr.........................ValAsnSer..... 510

1158 ACCCAGACTGAATAATGGAGAAAATATTTCCTTTATTGACGAAGGAAAAG 1207
           :::|||   |||:::|||       :::|||    ::::::||||
 511 ....SerIleAsnIleGlySerAsnSerHisLeuIleLeuHisSerLysG 526

1208 GCGAATTGATACTTACCAGCAACATCAATCAAGGTGCTGGAGGATTATAT 1257
     ||                            |||    :::|||||||:::
 526 ly........................GlnArgGlyGlyGlyValGln 533

1258 TTCCAAGGAGATTTTACGGTCTCGCCTGAAAATAACGAAACTTGGCAAGG 1307
     ::::::|||||||:::|||    :::       |||          :::::||
 534 IleAspGlyAspIleThrSerLysGlyGlyAsnLeuThrIleTyrSerGl 550

1308 CGCGGGCGTTCATATCAGTGAAGACAGTACCGTTACTTGGAAAGTAAACG 1357
     |:::    |||    :::    ::::::    |||:::
 550 yGlyTrpValAspValHisLysAsnIleThrLeu................ 561

1358 GCGTGGCAAACGACCGCCTGTCCAAAATCGGCAAAGGCACGCTGCACGTT 1407

 561 .................................................. 561

1408 CAAGCCAAAGGGGAAAACCAAGGCTCGATCAGCGTGGGCGACGGTACAGT 1457
                   :::||||||    :::::::::       ::::::||
 562 ..............AspGlnGlyPheLeuAsnIleThrAlaAlaSerVa 573

1458 CATTTTGGATCAGCAGGCAGACGATAAAGGCAAAAAACAAGCCTTTAGTG 1507
     |    :::    :::    :::::::::|||::::::       |||    ::::
 573 lAlaPhe...GluGlyGlyAsnAsnLysAlaArgAspAlaAlaAsnAlaL 589

1508 AAATCGGCTTGGTCAGCGGCAGGGGTACGGTGCAACTGAATGCCGATAAT 1557
     ::|||           ::::::|||||||||    ::::::::::::::
 589 ysIle.........ValAlaGlnGlyThrValThrIleThrGlyGluGly 602

1558 CAGTTCAACCCCGACAAACTCTATTTCGGCTTTCGCGGCGGACGTTTGGA 1607
     :::                        ||||||::::::::::::::
 603 Lys......................AspPheArgAlaAsnAsnValSe 611

1608 TTTAAACGGGCATTCGCTTTCGTTCCACCGTATTCAAAATACCGATGAAG 1657
     :|||||||||                                |||    :::|
 611 rLeuAsnGly.............................ThrGlyLysG 618

1658 GGGCGATGATTGTC............AACCACAATCAAGACAAAGAATCC 1695
     ||       |||:::             |||    :::::::::    ::::::
 618 lyLeuAsnIleIleSerSerValAsnAsnLeuThrHisAsnLeuSerGly 634

1696 ACCGTTACCATTACAGGCAATAAAGATATTGCTACAACCGGCAATAACAA 1745
     |||::::::|||:::||||||       |||       |||
 635 ThrIleAsnIleSerGlyAsnIleThrIleAsnGlnThr........... 647

1746 CAGCTTGGATAGCAAAAAAGAAATTGCCTACAACGGTTGGTTTGGCGAGA 1795
                ::::::|||:::     :::|||    |||       :::
 648 .........ThrArgLysAsnThrSerTyr......TrpGlnThrSerH 659

1796 AAGATACG.......................ACCAAAACGAACGGGCGG 1821
       |||:::                       :::|||:::::::::
 659 isAspSerHisTrpAsnValSerAlaLeuAsnLeuGluThrGlyAlaAsn 675

1822 CTCAACCTTGTTTACCAGCCCGCCGCAGAAGACCGCACCCTGCTGCTT.. 1869
     ::::::::::::               ::::::::::::::    |||
 676 PheThrPheIleLysTyrIleSerSerAsnSerLysGlyLeuThrThrGl 692
```

```
1870 .........TCCGGCGGAACAAATTTAAACGGCAACATCACGCAAACAA 1909
               |||:::|||:::|||:::||||||           :::|
 692 nTyrArgSerSerAlaGlyValAsnPheAsnGly...........ValA 705

1910 ACGGCAAACTGTTTTCAGCGGCAGACCAACACCGCACGCCTACAATCAT 1959
      |||||:::::   |||
 705 snGlyAsnMetSerPhe................................ 710

1960 TTAAACGACCATTGGTCGCAAAAAGAGGGCATTCCTCGCGGGGAAATCGT 2009
                    :::  |||||||||      :::       ::
 711 ..............AsnLeuLysGluGlyAlaLysValAsnPheLysLe 722

2010 GTGGGACAACGACTGGATCAACCGCACA.................TTTA 2041
     :       |||:::   :::|||   :::                 |||
 722 uLysProAsnGluAsnMetAsnThrSerLysProLeuProIleArgPheL 739

2042 AAGCGGAAAACTTCCAAATTAAAGGCGGACAGGCGGTGGTT......... 2082
       |||   |||:::         ||||||    :::|||
 739 euAla...AsnIleThrAlaThrGlyGlyGlySerValPhePheAspIle 754

2083 ...........TCCCGCAATGTTGCCAAAGTGAAA...GGCGATTGGCA 2117
                |||    :::   |||:::::|||   ::::::   ::
 755 TyrAlaAsnHisSerGlyArgGlyAlaGluLeuLysMetSerGluIleAs 771

2118 TTTGAGCAATCACGCCCAAGCAGTTTTTGGTGTCGCACCGCATCAAAGCC 2167
     :::::||||||   |||:::   ::::::               :::||||
 771 nIleSerAsnGlyAlaAsnPheThrLeu..............AsnSerH 783

2168 ACACAATCTGTACACGTTCGGACTGGACGGGTCTGACAAATTGTGTCGAA 2217
      ||          :::|||:::|||    :::          |||     :::
 783 is.........ValArgGlyAspAspAlaPheLysIleAsn...LysAsp 795

2218 AAAACCATTACCGACGATAAAGTGATTGCTTCATTGACTAAGACCGACAT 2267
           ||||||:::        :::        ||||||    :::|||
 796 LeuThrIleAsnAlaThrAsnSerAsnPheSerLeuArgGlnThrLysAs 812

2268 CAGCGGCAATGTCGATCTTGCCGATCACGCTCATTTAAATCTCACAGGGC 2317
     :::                |||    :::|||        :::        ::::
 812 pAspPheTyrAspGlyTyrAlaArgAsnAlaIleAsnSerThrTyrAsnI 829

2318 TTGCCACACTCAACGGCAATCTTAGTGCAAATGGCGATACACGTTATACA 2367
      :::::    |||:::|||||||::::::    :::|||
 829 leSerIleLeuGlyGlyAsnValThrLeuGlyGly.............. 840

2368 GTCAGCCACAACGCCACCCAAAACGGCAACCTTAGCCTCGTGGGCAATGC 2417
                        ||||||::::::    |||:::::||||||
 841 .................GlnAsnSerSerSerSerIleThrGlyAsn.. 850

2418 CCAAGCAACATTTAATCAAGCC.........ACATTAAACGGCAACACAT 2458
          |||:::::::::|||          ||||||:::::|||::::
 851 ....IleThrIleGluLysAlaAlaAsnValThrLeuGluAlaAsnAsnA 866

2459 CGGCTTCGGGCAATGCTTCATTTAATCTAAGCGACCACGCCGTACAAAAC 2508
      ::    :::    :::            |||:::    |||::::::::::::
 866 laProAsnGlnGln.........AsnIleArgAspArgValIleLysLeu 879

2509 GGCAGTCTGACGCTTTCCGGCAACGCTAAGGCAAACGTAAGCCATTCCGC 2558
      |||||||||   :::::::|||:::    :::    :::    :::::::::
 880 GlySerLeuLeuValAsnGlySerLeuSerLeuThrGlyGluAsnAlaAs 896

2559 ACTCAACGGTAATGTCTCCCTAGCCGATAAGGCAGTATTCCAT....... 2601
      :::::::||||||:::::::::::::::::::::::|||:::|||
 896 pIleLysGlyAsnLeuThrIleSerGluSerAlaThrPheLysGlyLysT 913

2602 ..TTTGAAAGCAGCCGCTTTACCGGACAAATCAGCGGCGGCAAGGATACG 2649
          ::::::    ::::::||||||:::::::::::::::       :::
 913 hrArgAspThrLeuAsnIleThrGlyAsnPheThrAsnAsnGlyThrAla 929

2650 GCATTACACTTAAAAGACAGCGAATGGACGCTGCCGTCAGGCACGGAATT 2699
        :::::::::    ::::::            |||    :::    |||:::
 930 GluIleAsnIleThrGlnGlyValValLysLeuGlyAsnValThrAsnAs 946

2700 AGGCAATTTAAACCTTGACAACGCCACCATTACACTCAATTCCGCCTATC 2749
      |||:::||||||:::    :::
 946 pGlyAspLeuAsnIleThrThr........................... 953

2750 GCCACGATGCGGCAGGGGCGCAAACCGGCAGTGCGACAGATGCGCCGCGC 2799

 953 .................................................. 953

2800 CGCCGTTCGCGCCGTTCGCGCCGTTCCCTATTATCCGT....TACACCGC 2845
        :::::::::|||::::::|||||||:::::::::
 954 ...HisAlaLysArgAsnGlnArgSerIleIleGlyGlyAspIleIleAs 969

2846 CAACTTCGGTAGAATCCCGTTTCAACACGCTGACGGTAAACGGCAAATTG 2895
      :::    |||    :::    :::          :::::::::      ::::|||:
 969 nLysLysGlySerLeuAsnIleThrAspSerAsnAsnAspAlaGluIleG 986

2896 AACGGTCAGGGAACATTCCGCTTTATGTCGGAACTCTTCGGCTACCGCAG 2945
      ::    :::|||||||||              |||:::|||    :::    ||
 986 lnIleGlyGlyAsnIleSerGlnLysGluGlyAsnLeuThrIle.SerSe 1002

2946 CGACAAATTGAAGCTGGCGGAAAGTTCCGAAGGCACTTACACCTTGGCGG 2995
      ||||||::::::::::::::
1002 rAspLysIleAsnIleThr.............................. 1008

2996 TCAACAATACCGGCAACGAACCTGCAAGCCTCGAACAATTGACGGTAGTG 3045
                                     :::|||:::|||:::
1009 ..............................LysGlnIleThrIleLys 1014

3046 GAAGGAAAAGACAACAAACCGCTGTCC...........GAAAACCTTAA 3083
      :::|||    |||::::::          |||          :::|||    ||
1015 LysGlyIleAspGlyGluAspSerSerSerAspAlaThrSerAsnAlaAs 1031

3084 TTTCACCCTGCAAAACGAACACGTCGAT.....................G 3112
      |:::|||::::::::::::::::::                              :
1031 nLeuThrIleLysThrLysGluLeuLysLeuThrGluAspLeuSerIleS 1048

3113 CCGGCGCGTGGCGTTACCAACTCATCCGCAAAGACGGC...GAGTTCCGC 3159
      ::|||        :::    ::::::          |||||||||    ::::::
1048 erGlyPheAsnLysAlaGluIleThrAlaLysAspGlyArgAspLeuThr 1064

3160 CTGCATAATCCGGTCAAAGAACAAGAGCTTTCCGACAAACTCGGCAAGGC 3209
      :::    |||                      :::|||        :::    ||
1065 IleGlyAsnSer.................AsnAspGlyAsnSerGlyAl 1075

3210 AGAAGCCAAAAAACAGGCGGAAAAAGACAACGCGCAAAGCCTTGACGCGC 3259
      |||||||    |||    :::          :::|||::::::        |||:::
1075 aGluAla...LysThrValThrPheAsnAsnValLys......AspSerL 1089

3260 TGATTGCGGCCGGGCGCGATGCCGTCGAAAAGACAGAAAGCGTTGCCGAA 3309
        |||:::|||              |||          :::::::::|||
1089 ysIleSerAlaAspGlyHisAsnValThrLeuAsnSerLysValLysThr 1105

3310 CCGGCCCGGCAGGCAGGCGGGGAAAATGTCGGCATTATGCAGGCGGAGGA 3359
        :::    :::::::|||    |||:::
1106 SerSerSerAsnGlyGlyArgGluSer...................... 1114

3360 AGAGAAAAAACGGGTGCAGGCGGATAAAGACACCGCCTTGGCGAAACAGC 3409
                      ::::::|||:::|||||||:::|||
1115 ...............AsnSerAspAsnAspThrGlyLeu.......... 1122

3410 GCGAAGCGGAAACCCGGCCGGCTACCACCGCCTTCCCCCGCGCCCGCCGC 3459
                                  |||          :::          :::
1123 ........................ThrIleThrAlaLysAsnValGlu 1130

3460 GCCCGCCGGGATTTGCCGCAACTGCAACCCCAACCGCAGCCCCAACCGCA 3509
      :::::::::|||:::    :::|||
1131 ValAsnLysAspIleThrSerLeu......................... 1138

3510 GCGCGACCTGATCAGCCGTTATGCCAATAGCGGTTTGAGTGAATTTTCCG 3559
                      :::    :::|||          ||||||          :
1139 ...............LysThrValAsnIleThrAlaSerGluLysValT 1150

3560 CCACGCTCAACAGCGTTTTCGCCGTACAGGACGAATTAGACCGCGTATTT 3609
```

```
          ::||| :::::::::   :::            ||||||      |||   ::
    1150 hrThr.ThrAlaGlySerThrIleAsnAlaThrAsnGlyLysAlaSerIl 1166

    3610 GCCGAAGACCGCCGCAACGCCGTTTGGACAAGCGGCATCCGGGACACCAA 3659
         :          ::::::         |||      :::::::::|||   :
    1166 eThrThrLysThrGlyAspIleSerGlyThrIleSerGlyAsnThrValS 1183

    3660 ACACTACCGTTCGC...AAGATTTCCGCGCCTACCGCCAACAAACCGACC 3706
         ::::::::::            ::::::::   :::::::::|||
    1183 erValSerAlaThrGlyAspLeuThrThrLysSerGlySerLysIleGlu 1199

    3707 TGCGCCAAATCGGTATGCAGAAAAACCTCGGCAGCGGGCGCGTCGGCATC 3756
         :::   ::::::   ::::::   |||||||||:::|||:::    :::::
    1200 AlaLysSerGlyGluAlaAsnValThrSerAlaThrGlyThrIleGlyGl 1216

    3757 CTGTTTTCGCACAACCGGACCGAAAACACCTTCGACGACGGCATCGGCA. 3805
         :           |||   :::   :::|||   ::::::|||||||:::|||
    1216 y.........ThrIleSerGlyAsnThrValAsnValThrAlaAsnAlaG 1230

    3806 ..................................ACTCGGCACGGCTTG 3820
                                           :::    :::|||
    1230 lyAspLeuThrValGlyAsnGlyAlaGluIleAsnAlaThrGluGlyAla 1246

    3821 CCCACGGCGCCGTTTTCGGGCAATACGGCATCGACA......GGTTCTAC 3864
              |||          :::|||||||||    :::|||        ||||||::
    1247 AlaThrLeuThrAlaThrGlyAsnThrLeuThrThrGluAlaGlySerSe 1263

    3865 ATCGGCATCAGCGCGGGCGCGGGTTTAGCAGCGGCAGCCTTTCAGACGG 3914
         :    :::|||::::::::::::|||           |||      |||::::
    1263 rIleThrSerThrLysGlyGlnValAspLeuLeuAla......GlnAsnG 1278

    3915 CATCGGAGGCAAAATCCGCCGCCGCGTGCTGCATTACGGCATTCAGGCAC 3964
         ::|||    |||    |||         |||:::     :::|||
    1278 lySerIleAlaGlySerIleAsnAlaAlaAsnValThrLeuAsnThrThr 1294

    3965 GATACCGCGCCGGTTTCGGCGGATTCGGCATCGAACCGCACATCGGCGCA 4014
            |||        :::   |||   |||        :::     ||||||:::
    1295 GlyThrLeuThrThrValAlaGlySerAspIleLysAlaThrSerGlyTh 1311

    4015 ACGCGCTATTTCGTCCAAA 4033
              :::|||:::::|||
    1311 rLeuValIleAsnAlaLys 1317
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01848

```
seq_documentation_block:
ID   AAB01848 standard; Protein; 1477 AA.
XX
AC   AAB01848;
XX
DT   11-SEP-2000  (first entry)
XX
DE   Haemophilus influenzae strain 12 HMW2A protein, SEQ ID NO:71.
XX
KW   HMW protein; hmw gene; hmwA1; hmwA2; high molecular weight;
KW   non-typeable Haemophilus influenzae; NTHi; non-encapsulated;
KW   recombinant production; Escherichia coli; antibacterial; vaccine;
KW   human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;
KW   detection; diagnosis.
XX
OS   Haemophilus influenzae strain 12.
XX
PN   WO200020609-A2.
XX
PD   13-APR-2000.
XX
PF   07-OCT-1999;    99WO-CA00938.
XX
PR   07-OCT-1998;    98US-0167568.
PR   08-DEC-1998;    98US-0206942.
XX
PA   (CONN-) CONNAUGHT LAB LTD.
XX
PI   Loosmore SM,  Yang Y,  Klein MH;
XX
DR   WPI; 2000-303789/26.
DR   N-PSDB; AAA52197.
XX
PT   Nucleic acid molecule for producing recombinant high molecular weight
PT   proteins of Haemophilus which are used as a vaccine to provide
PT   protection against Haemophilus induced diseases in humans -
XX
PS   Example 16; Fig 29A-N; 307pp; English.
XX
CC   The invention relates to the recombinant production of Haemophilus
CC   influenzae high molecular weight (HMW) proteins in Escherichia coli. The
CC   expression construct used to effect recombinant expression comprises a
CC   promoter functional in E. coli (e.g., the T7 promoter) operably linked
CC   to a modified hmwABC operon from a non-typeable (non-encapsulated) H.
CC   influenzae (NTHi). Most HMW-expressing NTHi strains contain two hmw gene
CC   clusters termed hmw1ABC and hmw2ABC. Each hmwABC operon comprises hmwA,
CC   hmwB and hmwC genes. The hmwA genes encode the structural HMWA proteins
CC   and the hmwB and hmwC genes encode accessory proteins which are
CC   responsible for post-translational processing and secretion of the HMWA
CC   proteins. The modified hmwABC operon used in the expression construct of
CC   the invention contains an A gene modified such that it encodes only the
CC   mature HMWA. The invention also discloses hmwA genes (AAA52175-A52198)
CC   and HMWA proteins (AAB01824-B01849) from the non-typeable H. influenzae
CC   strains Joyc, K1, K21, LCDC2, PMH1, 15 and 12. The nucleic acids and
CC   vectors are used for the production of recombinant H. influenzae HMW
CC   proteins which can be used as vaccines to mediate a humoral or
CC   cell-mediated immune response to provide protection against diseases in
CC   humans caused by H. influenzae (e.g., otitis media, epiglottitis,
CC   pneumonia and tracheobronchitis). The HMW proteins are also useful as
CC   antigens in immunoassays for detecting antibodies against Haemophilus,
CC   HMW proteins and/or HMW peptides. The nucleotide sequences encoding the
CC   HMW proteins can be used to isolate and clone hmw genes from other
CC   non-typeable strains of Haemophilus via hybridisation reactions. The
CC   present sequence represents an HMWA protein from a non-typeable strain of
CC   H. influenzae.
XX
SQ   Sequence   1477 AA;

alignment_scores:
           Quality:  285.00             Length:   1441
             Ratio:  0.417                Gaps:     67
 Percent Similarity:  47.398    Percent Identity: 19.778

alignment_block:
 US-09-303-518D-649 x AAB01848  ..

 Align seg 1/1  to: AAB01848  from: 1  to: 1477

     163 CGCGACTTTGCCGAAAATAAAGGCAAGTTTGCAGTCGGGGCGAAAGATAT 212
          :::|||         |||   :::   |||::::::::::::   |||||
     158 LysAspAlaIleIleAsnThrAsnGlyPheThrAlaSerThrLeuAspIl 174

     213 TGAGGTTTACAACAAAAAAGGGGAGTTGGTCGGCAAATCAATGACAAAA. 261
          |:::        |||   |||::::::            ::::::||||||
     174 eSerAsnGluAsnIleLysAlaArgAsnPheThrPheGluGlnThrLysA 191

     262 ..........GCCCCGATGATTGATTTTTCTGTG......GTGTCGCGT 294
                   |||   :::::::::   ::::::      |||::::::
     191 spLysAlaLeuAlaGluIleValAsnHisGlyLeuIleThrValGlyLys 207

     295 AACGGCGTGGCGGCATTGGTGGGC...........GATCAATATATTGT 332
          :::|||   :::   |||:::|||             ::::::   :::::
     208 AspGlySerValAsnLeuIleGlyGlyLysValLysAsnGluGlyValIl 224

     333 GAGCGTGGCACATAACGGCGGCTATAACAACGTTGATTTTGGT....... 375
          :||||||      |||||||||      ::::::      |||
     224 eSerVal......AsnGlyGlySerIleSerLeuLeuAlaGlyGlnLysI 239

     376 ........GCGGAAGGAAGAAATCCCGATCAACATCGTTTTACTTATAAA 417
                 ::::::        ||||||          :::||||||:::
```

```
 239 leThrIleSerAspIleIleAsnPro.........ThrIleThrTyrSer 252

 418 ATTGTGAAACGGAATAATTATAAAGCAGGGACTAAAGGCCATCCTTAT.. 465
     |||:::     :::|||   :::|||   :::   |||     :::
 253 IleAlaAlaProGluAsn...GluAlaValAsnLeuGlyAspIlePheAl 268

 466 ....GGCGGCGATTATCATATGCCGCGTTTGCATAAATTTGTCACAGATG 511
         ||||||:::   :::
 268 aLysGlyGlyAsnIleAsn............................... 274

 512 CAGAACCTGTTGAAATGACCAGTTATATGGATGGGCGGAAATATATCGAT 561

 274 .................................................. 274

 562 CAAAATAATTACCCTGACCGTGTTCGTATTGGGGCAGGCAGGCAATATTG 611
                           ||||||   ::::::   |||:::
 275 .....................ValArgAlaAlaThrIleArgAsnGlnGl 284

 612 GCGATCTGATGAAGATGAGCCCAATAACCGCGAAAGTTCATATCATATTG 661
      :::   :::   |||:::   ::::::   :::|||:::     :::
 284 yLysLeuSerAlaAspSerValSerLysAspLysSerGlyAsnIleValL 301

 662 CAAGTGCGTATTCTTGG......CTCGTTGGTGGCAATACCTTTGCACAA 705
       ||||||   :::               :::||||||        ||||||
 301 euSerAlaLysGluGlyGluAlaGluIleGlyGlyValIleSerAlaGln 317

 706 AAT............................................... 708
     |||
 318 AsnGlnGlnAlaLysGlyGlyLysLeuMetIleThrGlyAspLysValTh 334

 709 ...............................GGATCAGGTGGTGGCACAG 727
                                    |||:::   ||||||   :
 334 rLeuLysThrGlyAlaValIleAspLeuSerGlyLysGluGlyGlyGluT 351

 728 TCAACTTAGGTAGTGAAAAAATTAAACATAGCCCATATGGTTTT...... 771
     ::   ||||||:::::::::     ::::::     |||:::
 351 hrTyrLeuGlyGlyAspGluArgGlyGluGlyLysAsnGlyIleGlnLeu 367

 772 ...............TTACCAACAGGAGGCTCATTTGGCGACAGTGGCTC 806
                     |||     |||::::::::::::   ||||||
 368 AlaLysLysThrSerLeuGluLysGlySerThrIleAsnValSerGly.. 383

 807 ACCAATGTTTATCTATGATGCCCAAAAGCAAAAG................ 840
                                |||::::|||
 384 ........................LysGluLysGlyGlyArgAlaIleV 392

 841 ..TGG............TTAATTAATGGGGTATTGCAAACG...GGCAAC 873
       |||             ||||||:::|||   :::::::::   |||:::
 392 alTrpGlyAspIleAlaLeuIleAspGlyAsnIleAsnAlaGlnGlySer 408

 874 CCCTATATAGGAAAAAGCAATGGCTTC....................... 900
         |||:::|||::::::||||||
 409 GlyAspIleAlaLysThrGlyGlyPheValGluThrSerGlyHisTyrLe 425

 901 ................CAGCTGGTTCGT...AAAGATTGGTTCTATGATG 931
                     :::|||:::   |||:::|||:::   |||
 425 uSerIleAspSerAsnAlaIleValLysThrLysGluTrpLeuLeuAspP 442

 932 AAATCTTTGCTGGAGATACCCATTCAGTATTCTACGAACCACGTCAAAAT 981
                   |||     ::::::     |||          :::
 442 ro...........AspAspValThrIleGluAlaGluAspProLeuArg 454

 982 GGGAAATACTCTTTTAACGACGAT...AATAATGGCACAGGAAAAATCAA 1028
     ::::::   ::::::||||||:::     :::||||||||||:::   ::
 455 AsnAsnThrGlyIleAsnAspGluPheProThrGlyThrGlyGluAlaSe 471

1029 TGCCAAACATGAACACAATTCTCTGCCTAATAGATTAAAAACACGAACCG 1078
     :           ::::::::::::|||   :::   |||   :::   |||:
 471 rAspProLysLysAsnSerGluLeuLysThrThrLeuThrAsnThrThrI 488

1079 TTCAATTGTTT.....................AATGTTTCTTTATCCGAG 1107
     :::::   :::                     |||::::::   |||:::
 488 leSerAsnTyrLeuLysAsnAlaTrpThrMetAsnIleThrAlaSerArg 504
```

```
1108 ACAGCAAGAGAACCTGTTTATCATGCTGCAGGTGGTGTCAACAGTTATCG 1157
                                             |||||||||
 505 LysLeuThr..........................ValAsnSer..... 510

1158 ACCCAGACTGAATAATGGAGAAAATATTTCCTTTATTGACGAAGGAAAAG 1207
             :::|||   |||:::|||       :::|||   ::::::||||
 511 ....SerIleAsnIleGlySerAsnSerHisLeuIleLeuHisSerLysG 526

1208 GCGAATTGATACTTACCAGCAACATCAATCAAGGTGCTGGAGGATTATAT 1257
     ||                                |||   :::||||||:::
 526 ly.........................GlnArgGlyGlyGlyValGln 533

1258 TTCCAAGGAGATTTTACGGTCTCGCCTGAAAATAACGAAACTTGGCAAGG 1307
     :::::||||||:::|||    :::        |||          ::::::||
 534 IleAspGlyAspIleThrSerLysGlyGlyAsnLeuThrIleTyrSerGl 550

1308 CGCGGGCGTTCATATCAGTGAAGACAGTACCGTTACTTGGAAAGTAAACG 1357
     |:::   |||   :::   ::::::   |||:::
 550 yGlyTrpValAspValHisLysAsnIleThrLeu................ 561

1358 GCGTGGCAAACGACCGCCTGTCCAAAATCGGCAAAGGCACGCTGCACGTT 1407

 561 .................................................. 561

1408 CAAGCCAAAGGGGAAAACCAAGGCTCGATCAGCGTGGGCGACGGTACAGT 1457
                   :::||||||   :::::::::        :::::::||
 562 ..............AspGlnGlyPheLeuAsnIleThrAlaAlaSerVa 573

1458 CATTTTGGATCAGCAGGCAGACGATAAAGGCAAAAAACAAGCCTTTAGTG 1507
     |   :::   :::   :::::::::|||::::::        |||   ::::
 573 lAlaPhe...GluGlyGlyAsnAsnLysAlaArgAspAlaAlaAsnAlaL 589

1508 AAATCGGCTTGGTCAGCGGCAGGGGTACGGTGCAACTGAATGCCGATAAT 1557
     ::|||                 ::::::|||||||||   :::::::::::::::
 589 ysIle.........ValAlaGlnGlyThrValThrIleThrGlyGluGly 602

1558 CAGTTCAACCCCGACAAACTCTATTTCGGCTTTCGCGGCGGACGTTTGGA 1607
                                        ||||||::::::::::::::
     :::
 603 Lys.......................AspPheArgAlaAsnAsnValSe 611

1608 TTTAAACGGGCATTCGCTTTCGTTCCACCGTATTCAAAATACCGATGAAG 1657
     :|||||||||                            |||     :::|
 611 rLeuAsnGly............................ThrGlyLysG 618

1658 GGGCGATGATTGTC............AACCACAATCAAGACAAAGAATCC 1695
     ||      |||:::            |||   :::::::::   ::::::
 618 lyLeuAsnIleIleSerSerValAsnAsnLeuThrHisAsnLeuSerGly 634

1696 ACCGTTACCATTACAGGCAATAAAGATATTGCTACAACCGGCAATAACAA 1745
     |||::::::|||:::||||||       |||        |||
 635 ThrIleAsnIleSerGlyAsnIleThrIleAsnGlnThr........... 647

1746 CAGCTTGGATAGCAAAAAAGAAATTGCCTACAACGGTTGGTTTGGCGAGA 1795
                   ::::::|||:::   :::|||     |||      :::
 648 ..........ThrArgLysAsnThrSerTyr......TrpGlnThrSerH 659

1796 AAGATACG.......................ACCAAAACGAACGGGCGG 1821
       |||:::                           :::|||:::::::::
 659 isAspSerHisTrpAsnValSerAlaLeuAsnLeuGluThrGlyAlaAsn 675

1822 CTCAACCTTGTTTACCAGCCCGCCGCAGAAGACCGCACCCTGCTGCTT.. 1869
     ::::::::::::            :::::::::::::::   |||
 676 PheThrPheIleLysTyrIleSerSerAsnSerLysGlyLeuThrThrGl 692

1870 .........TCCGGCGGAACAAATTTAAACGGCAACATCACGCAAACAA 1909
                 |||:::|||:::|||:::||||||         :::|
 692 nTyrArgSerSerAlaGlyValAsnPheAsnGly...........ValA 705

1910 ACGGCAAACTGTTTTCAGCGGCAGACCAACACCGCACGCCTACAATCAT 1959
     |||||::::::   |||
 705 snGlyAsnMetSerPhe................................ 710
```

```
1960 TTAAACGACCATTGGTCGCAAAAAGAGGGCATTCCTCGCGGGGAAATCGT 2009
                        :::  |||||||||         :::    ::
 711 ..............AsnLeuLysGluGlyAlaLysValAsnPheLysLe 722

2010 GTGGGACAACGACTGGATCAACCGCACA.................TTTA 2041
      :     |||:::  :::|||  :::                  |||
 722 uLysProAsnGluAsnMetAsnThrSerLysProLeuProIleArgPheL 739

2042 AAGCGGAAAACTTCCAAATTAAAGGCGGACAGGCGGTGGTT......... 2082
      |||  |||:::         ||||||   :::|||
 739 euAla...AsnIleThrAlaThrGlyGlyGlySerValPhePheAspIle 754

2083 ...........TCCCGCAATGTTGCCAAAGTGAAA...GGCGATTGGCA 2117
                |||   :::   |||::::::|||   ::::::   ::
 755 TyrAlaAsnHisSerGlyArgGlyAlaGluLeuLysMetSerGluIleAs 771

2118 TTTGAGCAATCACGCCCAAGCAGTTTTTGGTGTCGCACCGCATCAAAGCC 2167
      :::||||||   |||:::  ::::::               :::||||
 771 nIleSerAsnGlyAlaAsnPheThrLeu..............AsnSerH 783

2168 ACACAATCTGTACACGTTCGGACTGGACGGGTCTGACAAATTGTGTCGAA 2217
      ||        :::|||:::|||   :::          |||     :::
 783 is.........ValArgGlyAspAspAlaPheLysIleAsn...LysAsp 795

2218 AAAACCATTACCGACGATAAAGTGATTGCTTCATTGACTAAGACCGACAT 2267
        ||||||:::       :::           ||||||   :::|||
 796 LeuThrIleAsnAlaThrAsnSerAsnPheSerLeuArgGlnThrLysAs 812

2268 CAGCGGCAATGTCGATCTTGCCGATCACGCTCATTTAAATCTCACAGGGC 2317
      :::                  |||   :::|||      :::      ::::
 812 pAspPheTyrAspGlyTyrAlaArgAsnAlaIleAsnSerThrTyrAsnI 829

2318 TTGCCACACTCAACGGCAATCTTAGTGCAAATGGCGATACACGTTATACA 2367
      :::::   |||:::||||||:::::    :::|||
 829 leSerIleLeuGlyGlyAsnValThrLeuGlyGly.............. 840

2368 GTCAGCCACAACGCCACCCAAAACGGCAACCTTAGCCTCGTGGGCAATGC 2417
                        ||||||:::::   |||::::::||||||
 841 .................GlnAsnSerSerSerSerIleThrGlyAsn.. 850

2418 CCAAGCAACATTTAATCAAGCC.........ACATTAAACGGCAACACAT 2458
               |||::::::::::|||        ||||||::::::|||::::
 851 ....IleThrIleGluLysAlaAlaAsnValThrLeuGluAlaAsnAsnA 866

2459 CGGCTTCGGGCAATGCTTCATTTAATCTAAGCGACCACGCCGTACAAAAC 2508
      ::    :::    :::           |||:::   |||:::::::::::::
 866 laProAsnGlnGln.........AsnIleArgAspArgValIleLysLeu 879

2509 GGCAGTCTGACGCTTTCCGGCAACGCTAAGGCAAACGTAAGCCATTCCGC 2558
      |||||||||    ::::::|||:::    :::    :::    :::::::::
 880 GlySerLeuLeuValAsnGlySerLeuSerLeuThrGlyGluAsnAlaAs 896

2559 ACTCAACGGTAATGTCTCCCTAGCCGATAAGGCAGTATTCCAT........ 2601
      ::::::||||||::::::::::::::::::||||::::|||
 896 pIleLysGlyAsnLeuThrIleSerGluSerAlaThrPheLysGlyLysT 913

2602 ..TTTGAAAGCAGCCGCTTTACCGGACAAATCAGCGGCGGCAAGGATACG 2649
           ::::::    :::::||||||::::::::::::::      :::
 913 hrArgAspThrLeuAsnIleThrGlyAsnPheThrAsnAsnGlyThrAla 929

2650 GCATTACACTTAAAAGACAGCGAATGGACGCTGCCGTCAGGCACGGAATT 2699
        :::::::::   ::::::             |||   :::   |||:::
 930 GluIleAsnIleThrGlnGlyValValLysLeuGlyAsnValThrAsnAs 946

2700 AGGCAATTTAAACCTTGACAACGCCACCATTACACTCAATTCCGCCTATC 2749
      |||:::|||||||:::   :::
 946 pGlyAspLeuAsnIleThrThr........................... 953

2750 GCCACGATGCGGCAGGGGCGCAAACCGGCAGTGCGACAGATGCGCCGCGC 2799

 953 .................................................. 953

2800 CGCCGTTCGCGCCGTTCGCGCCGTTCCCTATTATCCGT....TACACCGC 2845
```

```
         :::::::::|||::::::||||||:::::::::
 954 ...HisAlaLysArgAsnGlnArgSerIleIleGlyGlyAspIleIleAs 969

2846 CAACTTCGGTAGAATCCCGTTTCAACACGCTGACGGTAAACGGCAAATTG 2895
      :::   |||   :::   :::         :::::::::     :::|||:
 969 nLysLysGlySerLeuAsnIleThrAspSerAsnAsnAspAlaGluIleG 986

2896 AACGGTCAGGGAACATTCCGCTTTATGTCGGAACTCTTCGGCTACCGCAG 2945
      ::    :::|||||||||             ||||:::|||   :::    ||
 986 lnIleGlyGlyAsnIleSerGlnLysGluGlyAsnLeuThrIle.SerSe 1002

2946 CGACAAATTGAAGCTGGCGGAAAGTTCCGAAGGCACTTACACCTTGGCGG 2995
      ||||||::::::::::::::
1002 rAspLysIleAsnIleThr............................. 1008

2996 TCAACAATACCGGCAACGAACCTGCAAGCCTCGAACAATTGACGGTAGTG 3045
                                        :::|||:::|||:::
1009 ..............................LysGlnIleThrIleLys 1014

3046 GAAGGAAAAGACAACAAACCGCTGTCC............GAAAACCTTAA 3083
      :::|||   |||::::::      |||             :::|||    ||
1015 LysGlyIleAspGlyGluAspSerSerSerAspAlaThrSerAsnAlaAs 1031

3084 TTTCACCCTGCAAAACGAACACGTCGAT.....................G 3112
      |:::|||::::::::::::::::::                       :
1031 nLeuThrIleLysThrLysGluLeuLysLeuThrGluAspLeuSerIleS 1048

3113 CCGGCGCGTGGCGTTACCAACTCATCCGCAAAGACGGC...GAGTTCCGC 3159
      :::|||    :::    ::::::         |||||||||    ::::::
1048 erGlyPheAsnLysAlaGluIleThrAlaLysAspGlyArgAspLeuThr 1064

3160 CTGCATAATCCGGTCAAAGAACAAGAGCTTTCCGACAAACTCGGCAAGGC 3209
      :::   |||                           :::|||     :::    ||
1065 IleGlyAsnSer.................AsnAspGlyAsnSerGlyAl 1075

3210 AGAAGCCAAAAAACAGGCGGAAAAAGACAACGCGCAAAGCCTTGACGCGC 3259
      ||||||    |||   :::         :::|||:::::       |||:::
1075 aGluAla...LysThrValThrPheAsnAsnValLys......AspSerL 1089

3260 TGATTGCGGCCGGGCGCGATGCCGTCGAAAAGACAGAAAGCGTTGCCGAA 3309
      |||:::|||                |||          :::::::::|||
1089 ysIleSerAlaAspGlyHisAsnValThrLeuAsnSerLysValLysThr 1105

3310 CCGGCCCGGCAGGCAGGCGGGGAAAATGTCGGCATTATGCAGGCGGAGGA 3359
      :::     ::::::|||    |||:::
1106 SerSerSerAsnGlyGlyArgGluSer...................... 1114

3360 AGAGAAAAAAACGGGTGCAGGCGGATAAAGACACCGCCTTGGCGAAACAGC 3409
                        ::::::|||:::||||||:::|||
1115 ................AsnSerAspAsnAspThrGlyLeu.......... 1122

3410 GCGAAGCGGAAACCCGGCCGGCTACCACCGCCTTCCCCCGCGCCCGCCGC 3459
                                  |||        :::      :::
1123 ........................ThrIleThrAlaLysAsnValGlu 1130

3460 GCCCGCCGGGATTTGCCGCAACTGCAACCCCAACCGCAGCCCCAACCGCA 3509
      :::::::::|||:::    :::|||
1131 ValAsnLysAspIleThrSerLeu......................... 1138

3510 GCGCGACCTGATCAGCCGTTATGCCAATAGCGGTTTGAGTGAATTTTCCG 3559
                        :::    :::|||            ||||||     :
1139 ................LysThrValAsnIleThrAlaSerGluLysValT 1150

3560 CCACGCTCAACAGCGTTTTCGCCGTACAGGACGAATTAGACCGCGTATTT 3609
      :::|||  :::::::::   :::              ||||||    |||    ::
1150 hrThr.ThrAlaGlySerThrIleAsnAlaThrAsnGlyLysAlaSerIl 1166

3610 GCCGAAGACCGCCGCAACGCCGTTTGGACAAGCGGCATCCGGGACACCAA 3659
      :         ::::::      |||       :::::::::|||     :
1166 eThrThrLysThrGlyAspIleSerGlyThrIleSerGlyAsnThrValS 1183

3660 ACACTACCGTTCGC...AAGATTTCCGCGCCTACCGCCAACAAACCGACC 3706
      :::::::::::         :::::::::    :::::::::|||
```

```
1183 erValSerAlaThrGlyAspLeuThrThrLysSerGlySerLysIleGlu 1199

3707 TGCGCCAAATCGGTATGCAGAAAAACCTCGGCAGCGGGCGCGTCGGCATC 3756
     :::   ::::::   ::::::   |||||||||:::|||:::   :::::
1200 AlaLysSerGlyGluAlaAsnValThrSerAlaThrGlyThrIleGlyGl 1216

3757 CTGTTTTCGCACAACCGGACCGAAAACACCTTCGACGACGGCATCGGCA. 3805
     :            |||   :::   :::|||   ::::::||||||:::|||
1216 y.........ThrIleSerGlyAsnThrValAsnValThrAlaAsnAlaG 1230

3806 ...................................ACTCGGCACGGCTTG 3820
                                        :::   :::|||
1230 lyAspLeuThrValGlyAsnGlyAlaGluIleAsnAlaThrGluGlyAla 1246

3821 CCCACGGCGCCGTTTTCGGGCAATACGGCATCGACA......GGTTCTAC 3864
        |||          :::|||||||||   :::|||      ||||||::
1247 AlaThrLeuThrAlaThrGlyAsnThrLeuThrThrGluAlaGlySerSe 1263

3865 ATCGGCATCAGCGCGGGCGCGGGTTTTAGCAGCGGCAGCCTTTCAGACGG 3914
     :   :::|||::::::::::::|||           |||      |||::::
1263 rIleThrSerThrLysGlyGlnValAspLeuLeuAla......GlnAsnG 1278

3915 CATCGGAGGCAAAATCCGCCGCCGCGTGCTGCATTACGGCATTCAGGCAC 3964
     ::|||   |||    |||       |||:::   :::|||
1278 lySerIleAlaGlySerIleAsnAlaAlaAsnValThrLeuAsnThrThr 1294

3965 GATACCGCGCCGGTTTCGGCGGATTCGGCATCGAACCGCACATCGGCGCA 4014
        |||       :::    |||   |||       :::    |||||||:::
1295 GlyThrLeuThrThrValAlaGlySerAspIleLysAlaThrSerGlyTh 1311

4015 ACGCGCTATTTCGTCCAAA 4033
        :::|||::::::|||
1311 rLeuValIleAsnAlaLys 1317
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:AAW30291

seq_documentation_block:

```
ID   AAW30291 standard; Protein; 1598 AA.
XX
AC   AAW30291;
XX
DT   14-APR-1998  (first entry)
XX
DE   Non-typeable Haemophilus high mol.wt. surface protein HMW3.
XX
KW   Non-typeable Haemophilus; high molecular weight surface protein;
KW   HMW3; immunogen; vaccine; otitis media.
XX
OS   Haemophilus influenzae strain 5.
XX
FH   Key              Location/Qualifiers
FT   Misc-difference  113
FT                    /note= "encoded by GTG"
FT   Misc-difference  864
FT                    /note= "encoded by TGT"
XX
PN   WO9736914-A1.
XX
PD   09-OCT-1997.
XX
PF   01-APR-1997;   97WO-US04707.
XX
PR   01-APR-1996;   96US-0617697.
XX
PA   (BARE/) BARENKAMP S J.
XX
PI   Barenkamp SJ;
XX
DR   WPI; 1997-503038/46.
DR   N-PSDB; AAT90992.
XX
PT   High molecular weight proteins of non-typeable Haemophilus
PT   influenzae - useful for vaccine production
XX
PS   Claim 1; Page 93-97; 183pp; English.
XX
CC   This protein comprises the high molecular weight surface protein
CC   HMW3 (125 kDa) of non-typeable Haemophilus influenzae strain 5 that
CC   has the immunological ability to protect against disease caused by
CC   a non-typeable Haemophilus strain and is characterised by at least
CC   one surface-exposed B-cell epitope that is recognised by monoclonal
CC   antibody AD6.  The HMW3 amino acid sequence was deduced from an
CC   isolated hmw3 gene (see AAT90992).  HMW1 (see AAW30293), HMW2 (see
CC   AAW30294) and HMW4 (see AAW30292) have also been identified.  A
CC   conjugate comprising HMW3 linked to an antigen, hapten or
CC   polysaccharide, and a synthetic peptide of 6-150 amino acids
CC   corresponding to at least protective epitope of HMW3 are also
CC   claimed.  HMW proteins, conjugates and peptides can be used in
CC   vaccines, as immunogens for preparation of antibodies and as
CC   antigens for detection of these antibodies.
XX
SQ   Sequence   1598 AA;
```

alignment_scores:

```
          Quality:  281.50           Length:   1497
            Ratio:  0.398              Gaps:     73
Percent Similarity:  47.295   Percent Identity: 20.574
```

alignment_block:
US-09-303-518D-649 x AAW30291 ..

Align seg 1/1 to: AAW30291 from: 1 to: 1598

```
163 CGCGACTTTGCCGAAAATAAAGGCAAGTTTGCAGTCGGGGCGAAAGATAT 212
    :::|||           |||   :::   |||::::::::::::   |||||
158 LysAspAlaIleIleAsnThrAsnGlyPheThrAlaSerThrLeuAspIl 174

213 TGAGGTTTACAACAAAAAAGGGGAGTTGGTCGGCAAATCAATGACAAAA. 261
    |:::         |||   |||:::::               ::::::||||||
174 eSerAsnGluAsnIleLysAlaArgAsnPheThrLeuGluGlnThrLysA 191

262 ...........GCCCCGATGATTGATTTTTCTGTG......GTGTCGCGT 294
               |||   :::::::::   ::::::      |||::::::
191 spLysAlaLeuAlaGluIleValAsnHisGlyLeuIleThrValGlyLys 207

295 AACGGCGTGGCGGCATTGGTGGGC............GATCAATATATTGT 332
    :::|||   :::   |||:::|||            ::::::   :::::
208 AspGlySerValAsnLeuIleGlyGlyLysValLysAsnGluGlyValIl 224

333 GAGCGTGGCACATAACGGCGGCTATAACAACGTTGATTTTGGTGCGGAAG 382
    :||||||         |||||||||      :::::::::  |||:::
224 eSerVal......AsnGlyGly......SerIleSer.LeuLeuAlaGly 236

383 GAAGAAATCCCGATCAACATCGTTTTACTTATAAAATTGTGAAACGGAAT 432
    ::::::|||   |||:::   ::::::            :::       :::||
237 GlnLysIleThrIleSerAspIleIleAsnProThrIleThrTyrSerIl 253

433 AATTATAAAGCAGGGACTAAAGGCCATCCTTATGGCGGCGATTATCATAT 482
    |         :::
253 eAlaAlaProGlu..................................... 257

483 GCCGCGTTTGCATAAATTTGTCACAGATGCAGAACCTGTTGAAATGACCA 532
       :::   :::|||||||||::::::
258 ..AsnGluAlaIleAsnLeuGlyAsp........................ 265

533 GTTATATGGATGGGCGGAAATATATCGATCAAAATAATTACCCTGACCGT 582
    ::::::         |||||||||||::::::::         |||:::
266 IlePheAlaLysGlyGlyAsnIleAsnValArgAlaAlaThrIleArg.. 281

583 GTTCGTATTGGGGCAGGCAGGCAATATTGGCGATCTGAT.GAAGATGAGC 631
                          |||   |||   |||        |||:::
282 .....................AsnLysGlyLysLeuSerAlaAspSerV 291

632 CCAATAACCGCGAAAGTTCATATCATATTGCAAGTGCGTATTCTTGG... 678
      ::::::   :::|||:::       :::   ||||||   :::
```

```
 291 alSerLysAspLysSerGlyAsnIleValLeuSerAlaLysGluGlyGlu 307

 679 ...CTCGTTGGTGGCAATACCTTTGCACAAAAT................. 708
        :::||||||          ||||||||
 308 AlaGluIleGlyGlyValIleSerAlaGlnAsnGlnGlnAlaLysGlyGl 324

 708 .................................................. 708

 324 yLysLeuMetIleThrGlyAspLysValThrLeuLysThrGlyAlaValI 341

 709 ..........GGATCAGGTGGTGGCACAGTCAACTTAGGTAGTGAAAAA 747
               |||:::  ||||||   :::  ||||||:::::::::
 341 leAspLeuSerGlyLysGluGlyGlyGluThrTyrLeuGlyGlyAspGlu 357

 748 ATTAAACATAGCCCATATGGTTTT.....................TTACC 776
        ::::::     |||:::                         |||
 358 ArgGlyGluGlyLysAsnGlyIleGlnLeuAlaLysLysThrThrLeuGl 374

 777 AACAGGAGGCTCATTTGGCGACAGTGGCTCACCAATGTTTATCTATGATG 826
       |||::::::::::::  ||||||
 374 uLysGlySerThrIleAsnValSerGly...................... 383

 827 CCCAAAAGCAAAAG..................TGG............TTA 846
       |||::::|||                 |||            |||
 384 .....LysGluLysGlyGlyArgAlaIleValTrpGlyAspIleAlaLeu 398

 847 ATTAATGGGGTATTGCAAACGGGCAACCCCTATATAGGAAAAAGCAATGG 896
     |||:::|||  :::::::::  :::      |||:::|||::::::||
 399 IleAspGlyAsnIleAsnAlaGlnGlyLysAspIleAlaLysThrGlyGl 415

 897 CTTC....................................CAGCTGG 907
     ||||                                       :::|
 415 yPheValGluThrSerGlyHisTyrLeuSerIleAspAspAsnAlaIleV 432

 908 TTCGT...AAAGATTGGTTCTATGATGAAATCTTTGCTGGAGATACCCAT 954
     ||:::   |||:::|||:::                 |||   :::
 432 alLysThrLysGluTrpLeuLeu.................AspProGlu 442

 955 TCAGTATTCTACGAACCACGTCAAAATGGGAAATACTCTTTTAACGACGA 1004
     :::|||       |||     :::   ::::::  ::::::::   ||
 443 AsnValThrIleGluAlaProSerAlaSerArgValGluLeuGlyAlaAs 459

1005 TAATAATGGCACAGGA............AAAATCAATGCCAAACATGAAC 1042
     |:::|||:::    :::            |||::::::  |||   ::::
 459 pArgAsnSerHisSerAlaGluValIleLysValThrLeuLysLysAsnA 476

1043 ACAATTCTCTGCCTAATAGATTAAAAACACGAACCGTTCAATTGTTTAAT 1092
     :::::||||||
 476 snThrSerLeu....................................... 479

1093 GTTTCTTTATCCGAGACAGCAAGAGAACCTGTTTATCATGCTGCAGGTGG 1142
     ::::::|||::::::|||:::   :::   :::        :::|||
 480 ThrThrLeuThrAsnThrThrIleSerAsnLeuLeuLysSerAlaHisVa 496

1143 TGTCAAC...AGTTATCGACCCAGACTGAATAATGGAGAAAATATTTCCT 1189
      ||||||   :::   |||   :::|||:::   :::::::::||||||
 496 lValAsnIleThrAlaArgArgLysLeuThrValAsnSerSerIleSer. 512

1190 TTATTGACGAAGGAAAAGGCGAATTGATACTTACCAGCAACATCAATCAA 1239
       |||::::::|||    ::::::|||||||||   |||:::   :::|||
 513 ..IleGluArgGly...SerHisLeuIleLeuHisSerGlu...GlyGln 526

1240 GGTGCTGGAGGATTATATTTCCAAGGAGATTTTACGGTCTCGCCTGAAAA 1289
     |||:::   |||:::   ::::::   |||:::|||   :::       ||
 527 GlyGlyGlnGlyValGlnIleAspLysAspIleThrSerGluGlyGlyAs 543

1290 TAACGAAACTTGGCAAGGCGCGGGCGTTCATATCAGTGAAGACAGTACC. 1338
     |          ::::::|||:::   |||   :::   ::::::   |||
 543 nLeuThrIleTyrSerGlyGlyTrpValAspValHisLysAsnIleThrL 560

1339 ....................GTTACTTGGAAAGTAAACGGCGTGGCAAAC 1368
                         :::|||   |||   :::   :::|||
 560 euGlySerGlyPheLeuAsnIleThrThrLysGluGlyAspIleAlaPhe 576

1369 GACCGCCTGTCCAAAATCGGCAAAGGCACGCTGCACGTTCAAGCCAAAGG 1418
     :::        :::|||   |||::::::::::|||   :::   |||
 577 Glu......AspLysSerGlyArgAsnAsnLeuThrIleThrAla..... 589

1419 GGAAAACCAAGGCTCGATCAGCGTGGGC......GACGGTACAGTCATTT 1462
              ||||||:::|||:::   |||         :::|||      :::
 590 .......GlnGlyThrIleThrSerGlyAsnSerAsnGlyPheArgPheA 604

1463 TGGATCAGCAGGCAGACGATAAAGGCAAAAAACAAGCCTTTAGTGAAATC 1512
      :::   :::   ::::::   |||   |||   :::|||::::::
 604 snAsnValSerLeuAsnSerLeuGlyGlyLysLeuSerPheThrAspSer 620

1513 GGCTTGGTCAGCGGCAGGGGTACGGTGCAACTGAATGCCGATAATCAGTT 1562
               ||||||   |||   :::   |||   :::|||:::||
 621 ArgGluAspArgGlyArgArgThr...LysGlyAsnIleSerAsnLysPh 636

1563 CAACCCCGACAAACTCTATTTCGGCTTTCGCGGCGGACGTTTGGATTTAA 1612
     |:::                                 |||   |||::::::
 636 eAsp............................GlyThrLeuAsnIleS 643

1613 ACGGGCAT.........TCGCTTTCG..................TTCCAC 1635
     ::|||             |||::::::                  |||:::
 643 erGlyThrValAspIleSerMetLysAlaProLysValSerTrpPheTyr 659

1636 CGTATTCAAAATACCGATGAAGGGGCGATGATTGTCAACCACAATCAAGA 1685
     |||             |||:::|||            |||
 660 Arg...........AspLysGlyArgThrTyrTrpAsn........... 668

1686 CAAAGAATCCACCGTTACCATTACAGGC............AATAAAGATA 1723
             :::|||:::::::::|||:::               |||   :::|
 669 ....ValThrThrLeuAsnValThrSerGlySerLysPheAsnLeuSerI 684

1724 TTGCTACAACCGGCAATAACAACAGCTTGGATAGCAAAAAAGAAATTGCC 1773
     ||    :::||||||::::::::::::         |||   ::::::
 684 leAspSerThrGlySerGlySerThrGlyProSerIleArgAsnAlaGlu 700

1774 TACAACGGTTGGTTTGGCGAGAAAGATACG..................AC 1805
        ||||||          :::|||   |||                  ::
 701 LeuAsnGlyIleThrPheAsnLysAlaThrPheAsnIleAlaGlnGlySe 717

1806 CAAAACGAACGGGCGGCTCAACCTTGTTTACCAGCCC............. 1842
     :   :::|||     ::::::         :::|||
 717 rThrAlaAsnPheSerIleLysAlaSerIleMetProPheLysSerAsnA 734

1843 ........GCCGCAGAAGACCGCACCCTGCTGCTTTCCGGCGGAACAAAT 1884
             |||      ::::::   :::   :::|||||||||   :::
 734 laAsnTyrAlaLeuPheAsnGluAspIleSerValSerGlyGlyGlySer 750

1885 TTAAACGGCAACATCACGCAAACAAACGGCAAACTGTTTTTCAGCGGCAG 1934
     ||||||     :::::::::   :::::::::::::::
 751 LeuAsnPheLysLeuAsnAlaSerSerSerAsnIle............. 762

1935 ACCAACACCGCACGCCTACAATCATTTAAACGACCATTGGTCGCAAAAAG 1984
         ||||||
 763 .GlnThrPro........................................ 765

1985 AGGGCATTCCTCGCGGGGAAATCGTGTGGGACAACGACTGGATCAACCGC 2034
                    |||   |||:::
 766 ..............GlyValIleIle........................ 769

2035 ACATTTAAAGCGGAAAACTTCCAAATTAAAGGCGGACAGGCGGTGGTTTC 2084
           |||::::::||||||::::::::::|||||||:::::::::
 770 ......LysSerGlnAsnPheAsnValSerGlyGlySerThrLeuAsnLe 784

2085 CCGCAATGTTGCCAAAGTGAAAGGCGATTGGCATTTGAGCAATCACGCCC 2134
      :::      ::::::::::::         :::   ::::::|||      :
 784 uLysAlaGluGlySerThrGluThrAlaPheSerIleGluAsnAspLeuA 801

2135 AAGCAGTTTTTGGTGTCGCACCGCATCAAAGCCACACAATCTGTACACGT 2184
      ::
 801 sn................................................ 801
```

```
2185 TCGGACTGGACGGGTCTGACAAATTGTGTCGAAAAAACCATTACCGACGA 2234
                                |||:::::    :::||||||
 802 .................LeuAsnAlaThrGlyGlyAsnIleThr...... 810

2235 TAAAGTGATTGCTTCATTGACTAAGACCGACATCAGCGGCAATGTCGATC 2284
            |||   ::::::    ||||||        |||      :
 811 .......IleArgGlnValGluGlyThrAspSerArgValAsnLysGlyV 825

2285 TTGCCGATCACGCTCATTTAAATCTCACAGGGCTTGCCACACTCAACGGC 2334
     ::|||          :::::::::::  |||              |||
 825 alAlaAlaLysLysAsnIleThrPheLysGly..............Gly 836

2335 AATCTTAGTGCAAATGGCGATACACGTTATACA...GTCAGCCACAACGC 2381
     |||:::::   :::::::::        |||   ::::::  |||::
 837 AsnIleThrPheGlySerGlnLysAlaThrThrGluIleLysGlyAsnVa 853

2382 CACCCAAAACGGCAACCTTAGCCTCGTGGGCAATGCCCAAGCAACATTTA 2431
      :|||   |||   |||                               :
 853 lThrIleAsnLysAsn..................................T 859

2432 ATCAAGCCACATTAAACGGCAACACATCGGCTTCGGGCAATGCTTCATTT 2481
     :::::||||||||:::|||     :::    |||:::::::::::    :::
 859 hrAsnAlaThrLeuArgGlyAlaAsnPheAlaGluAsnLysSerProLeu 875

2482 AATCTAAGCGACCACGCCGTACAAAACGGCAGTCTGACGCTTTCCGGCAA 2531
     |||:::::    ::::::::::||||||:::||||||   :::|||
 876 AsnIleAlaGlyAsnValIleAsnAsnGlyAsnLeuThrThrAlaGly.. 891

2532 CGCTAAGGCAAACGTAAGCCATTCCGCACTCAAC......GGTAATGTCT 2575
                                |||   :::|||     ||||||:::
 892 ....................SerIleIleAsnIleAlaGlyAsnLeuT 901

2576 CCCTAGCCGATAAGGCAGTATTCCATTTTGAAAGCAGCCGCTTTACCGGA 2625
     ::::::::       |||
 901 hrValSerLysGlyAla................................ 906

2626 CAAATCAGCGGCGGCAAGGATACGGCATTACACTTAAAAGACAGCGAATG 2675

 906 .................................................. 906

2676 GACGCTGCCGTCAGGCACGGAATTAGGCAATTTAAACCTTGACAACGCCA 2725
                                         ||||||:::     |
 907 .................................AsnLeuGlnAlaIleT 912

2726 CC...ATTACACTCAATTCCGCCTATCGCCACGATGCGGCAGGGGCGCAA 2772
      ||        |||:::|||   |||          |||       ||||||:::
 912 hrAsnTyrThrPheAsnValAlaGlySerPheAspAsnAsnGlyAlaSer 928

2773 ACCGGCAGTGCGACAGATGCGCCGCGCCGCCGTTCGCGCCGTTCGCGCCG 2822
     :::   |||   :::   :::       :::   :::      ::::::
 929 AsnIleSerIleAlaArgGlyGlyAlaLysPheLysAspIleAsnAsnTh 945

2823 TTCCCTATTATCCGTTACACCGCCAACTTCGGTAGAATCCCGTTTCAACA 2872
       |||   |||::::::|||          |||:::   ::::::   ::::::|
 945 rSerSerLeuAsnIleThr......ThrAsnSerAspThrThrTyrArgT 960

2873 CGCTGACGGTAAACGGCAAATTGAACGGTCAGGGAACATTCCGCTTTATG 2922
      ||:::      ::::::        |||   :::|||   ::::::::::::
 960 hrIleIleLysGlyAsnIleSerAsnLysSerGlyAspLeuAsnIleIle 976

2923 TCGGAACTCTTCGGCTACCGCAGCGACAAATTGAAGCTGGCGGAA..... 2967
                                 ||||||   :::   ||||||
 977 ........................AspLysLysSerAspAlaGluIleGl 985

2968 ..................AGTTCCGAAGGCACTTACACCTTGGCG.... 2994
                     ::::::||||||:::     |||::::::
 985 nIleGlyGlyAsnIleSerGlnLysGluGlyAsnLeuThrIleSerSerA 1002

2995 .....GTCAACAATACCGGCAACGAACCTGCAAGCCTCGAACAATTGACG 3039
          ||||||   |||                          :::|||:::|||
1002 spLysValAsnIleThr...................AsnGlnIleThr 1011

3040 GTA.........GTGGAA...GGAAAAGACAACAAACCGCTGTCCGAAAA 3077
      :::            ||||||     |||::::::::::::     :::||||||
1012 IleLysAlaGlyValGluGlyGlyArgSerAspSerSerGluAlaGluAs 1028

3078 CCTTAATTTCACCCTGCAAAACGAACACGTCGAT................ 3111
     |   |||:::|||:::|||::::::::::::
1028 nAlaAsnLeuThrIleGlnThrLysGluLeuLysLeuAlaGlyAspLeuA 1045

3112 .....GCCGGCGCGTGGCGTTACCAACTCATCCGCAAAGACGGC...GAG 3153
          :::|||       :::   ::::::        |||:::|||   :::
1045 snIleSerGlyPheAsnLysAlaGluIleThrAlaLysAsnGlySerAsp 1061

3154 TTCCGCCTGCATAATCCGGTCAAAGAACAAGAGCTTTCCGACAAACTCGG 3203
     :::   :::   |||                                :::   ||
1062 LeuThrIleGlyAsnAla.......................SerGlyGl 1070

3204 CAAGGCAGAAGCCAAAAAA.........CAGGCGGAAAAAGACAACGCGC 3244
     |:::|||:::|||||||||           :::      ||||||:::
1070 yAsnAlaAspAlaLysLysValThrPheAspLysValLysAspSerLysI 1087

3245 AAAGCCTTGACGCGCTGATTGCGGCCGGGCGCGATGCCGTCGAAAAGACA 3294
       |||   |||                |||::::::::::::   ::::::
1087 leSerThrAsp..............GlyHisAsnValThrLeuAsnSer 1098

3295 GAA...AGCGTTGCCGAACCGGCCCGGCAGGCAGGCGGGGAAAAT..... 3336
     |||   ::::::::::::   :::   :::||||||::::::|||
1099 GluValLysThrSerAsnGlySerSerAsnAlaGlyAsnAspAsnSerTh 1115

3337 .......................GTCGGCATTATGCAGGCGGAGGAAG 3361
                           |||   :::   :::            :
1115 rGlyLeuThrIleSerAlaLysAspValThrValAsnAsnAsnValThrS 1132

3362 AGAAAAAACGGGTGCAGGCGGATAAAGACACCGCCTTGGCGAAACAGCGC 3411
     ::   |||   ::::::   :::      ::::::   :::         :::
1132 erHisLysThrIleAsnIleSerAlaAlaAlaGlyAsnValThrThrLys 1148

3412 GAAGCGGAAACCCGGCCGGCTACCACCGCCTTCCCCCGCGCCCGCCGCGC 3461
     |||:::   |||         ||||||||||:::   |||         :::|||
1149 GluGlyThrThrIleAsnAlaThrThrGly.SerValGluValThrAlaG 1165

3462 CCGCCGGGATTTGCCGCAACTGCAACCCCAACCGCAGCCCCAACCGCAGC 3511
        |||      ::::::         :::::::::   :::   :::
1165 lnAsnGlyThrIleLysGlyAsnIleThrSerGlnAsnValThrValThr 1181

3512 GCGACCTGATCAGCCGTTATGCCAATAGCGGTTTGAGTGAATTTTCCGCC 3561
     ||||||   :::   |||                :::   :::|||
1182 AlaThrGluAsnLeuValThrThrGluAsnAlaValIleAsnAlaThrSe 1198

3562 ACGCTCAACAGCGTTTTCGCCGTACAGGACGAATTAGACCGCGTATTTGC 3611
      ::::::   :::|||
1198 rGlyThrValAsnIleSer............................T 1205

3612 CGAAGACCGCCGCAACGCCGTTTGGA...CAAGCGGCATCCGGGACACCA 3658
       ||||||:::                 |||   :::::::::||||||:::
1205 hrLysThrGlyAspIleLysGlyGlyIleGluSerThrSerGlyAsnVal 1221

3659 AACACTACCGTTCGC...........AAGATTTCCGCGCCTACCGCCAA 3696
     |||   |||:::               |||:::|||         |||::::
1222 AsnIleThrAlaSerGlyAsnThrLeuLysValSerAsnIleThrGlyGl 1238

3697 CAAACCGACCTGCGCCAAATCGGTATGCAGAAAAACCTCGGCAGCGGGCG 3746
     :       |||   :::   ::::::            |||::::::||||||:
1238 nAspValThrValThrAlaAspAlaGlyAlaLeuThrThrThrAlaGlyS 1255

3747 CGTCGGCATCCTGTTTTCGCACAACCGGA................... 3775
     :::::   |||:::      |||||||||
1255 erThrIleSerAla......ThrThrGlyAsnAlaAsnIleThrThrLys 1269

3776 ..........................CCGAAAACACCTTCGACGACGGC 3798
                                 ::::::   |||:::|||
1270 ThrGlyAspIleAsnGlyLysValGluSerSerSerGlySerValThrLe 1286

3799 ATCGGCAACTCGGCACGGCTTGCCCACGGCGCCGTTTTCGGGCAAT.... 3844
        ||||||          |||   ::::::   :::|||||||||
```

```
1286 uValAlaThrGlyAlaThrLeuAlaValGlyAsnIleSerGlyAsnThrV 1303

3845 ........ACGGCATCGACAGGTTCTACATCGGCATCAGCGCGGGCGCGG 3886
             |||||:::::|||:::   :::::::::::   :::
1303 alThrIleThrAlaAspSerGlyLysLeuThrSerThrValGlySerThr 1319

3887 GTTTTAGCAGCGGCAGCCTTTCAGACGGCATCGGAGGCAAAA........ 3928
      :::   ::::::   :::      |||:::|||:::::
1320 IleAsnGlyThrAsnSerValThrThrSerSerGlnSerGlyAspIleGl 1336

3929 ..........TCCGCCGCCGCGTGCTGCATTACGGCATTCAGGCACGAT. 3967
               |||:::   :::     :::||||||         |||
1336 uGlyThrIleSerGlyAsnThrValAsnValThrAlaSerThrGlyAspL 1353

3968 ..ACCGCGCCGGTTTCGGCGGATTCGGCATCGAACCGCACATCGGCGCAA 4015
       |||        |||||          :::::::::   :::|||
1353 euThrIleGlyAsnSerAlaLysValGluAlaLysAsnGlyAlaAlaThr 1369

4016 CGCGCTATTTCGTCCAAAAAGCGGATTACCGCTACGAAAACGTCAATATC 4065
        :::   :::|||   |||      |||   |||   :::|||||||::
1370 LeuThrAlaGluSerGlyLysLeuThrThrGlnThrGlySerSerIleTh 1386

4066 GCCACCCCCGGCCTTGCATTCAACCGCTACCGCGCGG 4102
      :          :::   ::::::|||   ||||||:::
1386 rSerSerAsnGly...GlnThrThrLeuThrAlaLys 1397
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:AAR41731

seq_documentation_block:

```
ID   AAR41731 standard; Protein; 1338 AA.
XX
AC   AAR41731;
XX
DT   26-APR-1994  (first entry)
XX
DE   High molecular weight protein 3 (HMW3).
XX
KW   HMW; high molecular weight protein; virus; vaccine; influenza;
KW   epitope; immunity; haemophilus influenzae.
XX
OS   Haemophilus influenzae.
XX
PN   WO9319090-A.
XX
PD   30-SEP-1993.
XX
PF   16-MAR-1993;   93WO-US02166.
XX
PR   16-MAR-1992;   92GB-0005704.
XX
PA   (BARE/) BARENKAMP S J.
PA   (INRM ) INSERM INST NAT SANTE & RECH  MEDICALE.
XX
PI   Barenkamp SJ;
XX
DR   WPI; 1993-320683/40.
DR   N-PSDB; AAQ49510.
XX
PT   High molecular weight surface proteins - of non-typeable
PT   haemophilus which exhibit immunogenic properties
XX
PS   Claim 5; Figure 10; 100pp; English.
XX
CC   The isolation and purification of the high molecular weight protein
CC   enables the identification of the major protective epitopes of the
CC   protein by conventional epitope mapping.  These epitopes can then be
CC   synthesised using standard techniques and incorporated into fully
CC   synthetic or recombinant vaccines.
XX
SQ   Sequence   1338 AA;
```

alignment_scores:

```
             Quality:  281.00          Length:   1478
               Ratio:  0.404             Gaps:     72
  Percent Similarity:  47.023   Percent Identity: 20.162
```

alignment_block:
 US-09-303-518D-649 x AAR41731 ..

 Align seg 1/1 to: AAR41731 from: 1 to: 1338

```
 127 GGACACACTTATTTCGGCATCAACTACCAATACTATCGCGACTTTGCCGA 176
     |||:::|||||||:::|||                         :::::
  89 GlyGluThrTyrLeuGly.........................GlyAs 96

 177 AAATAAAGGCAAGTTTGCAGTCGGGGCGAAAGATATTGAGGTTTACAACA 226
     :::::::|||:::          |||   :::   |||:::::   :::|
  96 pGluArgGlyGlu.........GlyLysAsnGlyIleGlnLeuAlaLysL 110

 227 AAAAAGGGGAGTTGGTCGGCAAATCAATGACAAAAGCCCCGATGATTGAT 276
     ||                  ::::::   |||:::      |||:::
 110 ys..................ThrThrLeuGluLysGlySerThrIleAsn 120

 277 TTTTCTGTGGTGTCGCGTAACGGCGTGGCGGCATTGGTGGGCGATCAATA 326
        |||      ::::::::|||   |||   :::   ||||||
 121 ValSerGlyLysGluLysGlyGlyArgAlaIleValTrpGlyAspIleAl 137

 327 TATTGTG..........................AGCGTGGCACATAACG 349
      ::::::                           ::::::|||   :::|
 137 aLeuIleAspGlyAsnIleAsnAlaGlnGlyLysAspIleAlaLysThrG 154

 350 GCGGCTAT.......................................... 357
     ||||||:::
 154 lyGlyPheValGluThrSerGlyHisTyrLeuSerIleAspAspAsnAla 170

 357 .................................................. 357

 171 IleValLysThrLysGluTrpLeuLeuAspProGluAsnValThrIleGl 187

 358 .............AACAACGTTGATTTTGGTGCGGAAGGAAGAAATCCCG 394
                  ::::::|||::::::|||||||:::   ||||||
 187 uAlaProSerAlaSerArgValGluLeuGlyAlaAsp...ArgAsnSerH 203

 395 ATCAACATCGTTTTACTTATAAAATTGTGAAACGGAATAATTATAAAGCA 444
       :::   :::              :::|||:::||||||   :::
 203 isSerAlaGluValIleLysValThrLeuLysLysAsnAsnThrSerLeu 219

 445 GGGACTAAAGGCCATCCTTATGGCGGCGATTATCATATGCCGCGTTTGCA 494
       |||       :::                           :::
 220 ThrThrLeuThrAsn.......................ThrThrIleSe 228

 495 TAAATTTGTCACAGATGCAGAACCTGTTGAAATGACCAGT.......... 534
      :::::::::   :::|||:::   |||::::::|||:::
 228 rAsnLeuLeuLysSerAlaHisValValAsnIleThrAlaArgArgLysL 245

 535 ..TATATGGATGGGCGGAAATATATCGATCAAAATAATTACCCTGACCGT 582
        :::::::::          |||:::::::::::::::
 245 euThrValAsnSerSerIleSerIleGluArgGlySerHisLeuIleLeu 261

 583 GTTCGTATTGGGGCAGGCAGGCAATATTGGCGATCTGATGAAGATGAGCC 632
              |||   |||   |||         :::   |||:::|||
 262 HisSerGluGlyGlnGlyGlyGlnGlyValGlnIleAspLysAsp..... 276

 633 CAATAACCGCGAAAGTTCATATCATATTGCAAGTGCGTATTCTTGGCTCG 682
                                 |||:::|||
 277 ........................IleThrSerGlu............. 280

 683 TTGGTGGCAATACCTTTGCACAAAATGGATCAGGTGGT............ 720
       |||||||||                  |||||||||
 281 ..GlyGlyAsnLeuThrIleTyr......SerGlyGlyTrpValAspVal 294

 721 ...GGCACAGTCAACTTAGGTAGTGAAAAAATTAAACATAGCCCATATGG 767
          :::::::::|||||||||                        ||
 295 HisLysAsnIleThrLeuGlySer.......................Gl 303
```

```
 768 TTTTTACCAACAGGAGGCTCATTTGGCGAC..................A 799
     ||||||           :::  |||||                    |
 303 yPheLeuAsnIleThrThrLysGluGlyAspIleAlaPheGluAspLysS 320

 800 GTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCAAAAGTGGTTAATT 849
     |||||         :::       |||||
 320 erGlyArgAsnAsnLeuThrIleThrAlaGln.................. 330

 850 AATGGGGTATTGCAAACGGGCAACCCCTATATAGGAAAAAGCAATGGCTT 899
        |||:::::   :::||||||              ||||||||||
 331 ...GlyThrIleThrSerGlyAsn..............SerAsnGlyPh 341

 900 CCAGCTGGTTCGTAAAGATTGGTTCTATGATGAAATCTTTGCTGGAGATA 949
     |:::                    ::::::::::::      ::::::
 341 eArg....................PheAsnAsnValSerLeuAsnSerL 351

 950 CCCATTCAGTATTCTACGAACCACGTCAAAATGGGAAATACTCTTTTAAC 999
                                   :::||||||   |||||||:::
 351 eu..........................GlyGlyLysLeuSerPheThr 358

1000 GACGATAAT......................................... 1008
     |||::::::
 359 AspSerArgGluAspArgGlyArgArgThrLysGlyAsnIleSerAsnLy 375

1009 ......................AATGGCACAGGAAAAATCAATGCCAAAC 1036
                           :::||||||     |||:::  |||
 375 sPheAspGlyThrLeuAsnIleSerGlyThrValAspIleSerMetLysA 392

1037 ATGAACACAATTCTCTGCCTAATAGATTAAAAACACGAACCGTT...... 1080
                |||        |||  |||  ||||||
 392 laProLysValSerTrpPheTyrArgAspLysGlyArgThrTyrTrpAsn 408

1081 ...........................CAATTGTTTAATGTTTCTTTATC 1103
                                :::   ||||||:::|||::::
 409 ValThrThrLeuAsnValThrSerGlySerLysPheAsnLeuSerIleAs 425

1104 CGAGACAGCAAGAGAACCTGTTTATCATGCTGCAGGTGGTGTCAACAGTT 1153
     ::::|||                                   |||  :::|||
 425 pSerThr..............................GlySerGlySerT 432

1154 ATCGACCCAGACTGAATAATGGAGAA......AATATTTCCTTT...... 1191
        |||   ::::::|||:::|||         :::|||:::|||
 432 hrGlyProSerIleArgAsnAlaGluLeuAsnGlyIleThrPheAsnLys 448

1192 ...........ATTGACGAAGGAAAAGGCGAATTGATACTTACCAGCAA 1229
                |||   :::|||:::              |||:::||
 449 AlaThrPheAsnIleAlaGlnGlySer..............ThrAlaAs 460

1230 CATCAATCAAGGTGCTGGAGGATTATATTTCCAAGGAGATTTTACGGTCT 1279
     |::::::      |||:::   :::   |||:::::::::  :::    :
 460 nPheSerIleLysAlaSerIleMetProPheLysSerAsnAlaAsnTyrA 477

1280 CGCCTGAAAATAACGAAACTTGGCAAGGCGCGGGCGTTCATATCAGTGAA 1329
     ::      |||::::::                    :::   :::|||
 477 laLeuPheAsnGluAsp..................IleSerValSerGly 487

1330 GACAGTACCGTTACTTGGAAAGTAAACGGC.................... 1359
        ::::::|||::::::|||:::|||:::
 488 GlyGlySerValAsnPheLysLeuAsnAlaSerSerSerAsnIleGlnTh 504

1360 ............GTGGCAAACGACCGCCTGTCCAAAATCGGCAAAGGCA 1396
                 :::   ::::::::::::::        |||    :::|
 504 rProGlyValIleIleLysSerGlnAsnPheAsnValSerGlyGlySerT 521

1397 CGCTGCACGTTCAAGCCAAAGGGGAAAACCAAGGCTCGATCAGCGTG... 1443
     |||||::::::::::|||:::|||:::::::::    ::::::|||:::
 521 hrLeuAsnLeuLysAlaGluGlySerThrGluThrAlaPheSerIleGlu 537

1444 ...........................GGCGACGGTACAGTCATTTTGGA 1466
                                |||:::   |||:::
 538 AsnAspLeuAsnLeuAsnAlaThrGlyGlyAsnIleThrIleArgGlnVa 554

1467 TCAGCAGGCAGACGATAAA............................... 1485
```

```
                 :::    :::|||::::::
 554 lGluGlyThrAspSerArgValAsnLysGlyValAlaAlaLysLysAsnI 571

1486 ............................GGCAAAAAACAAGCCTTTAGT 1506
                                 |||:::::::::|||     :::
 571 leThrPheLysGlyGlyAsnIleThrPheGlySerGlnLysAlaThrThr 587

1507 GAAATCGGCTTGGTCAGCGGCAGGGGTACGGTGCAACTGAATGCCGATAA 1556
     ||||||                      :::|||:::|||   :::|||   :::::
 588 GluIle..............LysGlyAsnValThrIleAsnLysAsnTh 599

1557 T.....................CAGTTCAACCCCGACAAACTCTATTTCG 1585
     :                       :::|||     :::|||      ::::
 599 rAsnAlaThrLeuArgGlyAlaAsnPheAlaGluAsnLysSerProLeuA 616

1586 GCTTTCGCGGCGGACGTTTGGATTTAAACGGGCATTCGCTTTCGTTCCAC 1635
     :::::  |||:::  ::::::  ||||||:::        :::
 616 snIleAlaGlyAsnValIleAsn...AsnGlyAsnLeuThrThrAlaGly 631

1636 CGTATTCAAAATACCGATGAAGGGGCGATGATTGTCAACCACAATCAAGA 1685
        |||   |||            |||   :::   |||:::   :::   ::
 632 SerIleIleAsnIle...AlaGlyAsnLeuThrValSerLysGlyAlaAs 647

1686 CAAAGAATCCACCGTTACCATTACAGGCAATAAAGATATTGCTACAACCG 1735
     :   ::::::  ::::::   |||   |||      :::|||   :::
 647 nLeuGlnAlaIleThrAsnTyrThrPheAsn......ValAlaGlySerP 662

1736 GCAATAACAACAGCTTGGATAGCAAAAAAGAAATTGCCTACAACGGTTGG 1785
        :::||||||:::         |||:::   :::||||||   :::|||
 662 heAspAsnAsnGlyAla...SerAsnIleSerIleAla...ArgGlyGly 676

1786 TTTGGCGAGAAAGATACGACCAAAACGAACGGGCGGCTCAACCTTGTTTA 1835
              ||||||   ::::::|||::::::      |||||::::::
 677 AlaLysPheLysAspIleAsnAsnThrSerSer...LeuAsnIleThrTh 692

1836 CCAGCCCGCCGCAGAAGACCGCACCCTGCTGCTTTCCGGCGGAACAAATT 1885
      :::        :::        ||||||:::                   :
 692 rAsnSerAspThrThrTyrArgThrIle.....................I 702

1886 TAAACGGCAACATCACGCAAACAAACGGCAAACTGTTTTTCAGCGGCAGA 1935
      ::::::||||||||||::::::   :::|||   |||   :::      :::
 702 leLysGlyAsnIleSerAsnLysSerGlyAspLeuAsnIleIleAspLys 718

1936 CCAACACCGCACGCCTACAATCATTTAAACGACCATTGGTCGCAAAAAGA 1985
        :::        |||      :::::::::   :::   ||||||||||
 719 LysSer...AspAlaGluIleGlnIleGlyGlyAsnIleSerGlnLysGl 734

1986 GGGCATTCCTCGCGGGGAAATCGTGTGGGACAACGACTGGATCAACCGCA 2035
     |                   |||:::::::::   ::::::|||   :::|||   |
 734 u...........GlyAsnLeuThrIleSerSerAspLysValAsnIleT 747

2036 CA...............................TTTAAAGCGGAAAAC 2052
     ||                                 ::::::::::::
 747 hrAsnGlnIleThrIleLysAlaGlyValGluGlyGlyArgSerAspSer 763

2053 TTCCAAATTAAAGGCGGACAGGCGGTGGTTTCCCGCAATGTTGCCAAAGT 2102
        :::   ::::::::::::   :::::::::   :::      |||::
 764 SerGluAlaGluAsnAlaAsnLeuThrIleGlnThrLysGluLeuLysLe 780

2103 GAAAGGCGATTGGCATTTGAGCAATCACGCCCAAGCAGTTTTTGGTGTCG 2152
     :   ||||||   ::::::|||:::         :::|||   :::   :::
 780 uAlaGlyAspLeuAsnIleSerGlyPheAsnLysAlaGluIleThrAlaL 797

2153 CACCGCATCAAAGCCACACAATCTGTACACGTTCGGACTGGACGGGTCTG 2202
              ::::::   ||||||   :::   |||
 797 ysAsnGlySerAspLeuThrIleGlyAsnAlaSerGly............ 809

2203 ACAAATTGTGTCGAAAAAACCATTACCGACGATAAAGTGATTGCTTCATT 2252
        |||:::      |||   :::|||   |||||||||      |||
 810 GlyAsnAlaAspAlaLysLysValThrPheAspLysValLysAspSerLy 826

2253 GACTAAGACCGACATCAGCGGCAATGTCGATCTTGCCGATCACGCTCATT 2302
        :::||||||   :::      ||||||   |||      ::::::
```

```
 826 sIleSerThrAsp...GlyHisAsnValThrLeuAsnSerGlu....... 839

2303 TAAATCTCACAGGGCTTGCCACACTCAACGGCAATCTTAGTGCAAATGGC 2352
                    :::  |||  ||||||:::  :::|||::::::
 840 ..............ValLysThrSerAsnGlySerSerAsnAlaGlyAsn 851

2353 GAT......ACACGTTATACAGTCAGC...CACAACGCCACCCAAAACGG 2393
     |||      |||     |||:::|||    ::::::|||  |||::
 852 AspAsnSerThrGlyLeuThrIleSerAlaLysAspValThrValAsnAs 868

2394 CAACCTTAGCCTCGTGGGCAATGCCCAAGCAACATTTAATCAAGCCACAT 2443
     :|||:::::              ::::::  |||:::|||  ::::::
 868 nAsnValThr............SerHisLysThrIleAsnIleSerAlaA 881

2444 TAAACGGCAACACATCGGCTTCGGGCAATGCTTCATTTAATCTAAGCGAC 2493
           ||||||:::::::::::  ::::::::::::|||   :::
 881 laAlaGlyAsnValThrThrLysGluGlyThrThrIleAsnAlaThrThr 897

2494 CACGCCGTA...........CAAAACGGCAGTCTGACGCTTTCCGGCAA 2531
        :::|||           |||||||||      |||:::::|||||
 898 GlySerValGluValThrAlaGlnAsnGly......ThrIleLysGlyAs 912

2532 C...GCTAAGGCAAACGTAAGCCATTCCGCACTCAACGGTAATGTCTCCC 2578
     |    ::::::  |||||||:::  :::|||  :::  |||:::
 912 nIleThrSerGlnAsnValThrValThrAlaThrGlu...AsnLeuValT 928

2579 TAGCCGATAAGGCAGTATTCCATTTGAAAGC..............AGC 2613
       ::::::::||||||::::::          |||            |||
 928 hrThrGluAsnAlaValIleAsnAlaThrSerGlyThrValAsnIleSer 944

2614 CGCTTTACCGGACAAATCAGCGGCGGCAAGGATACGGCA.........TT 2654
            ||||||:::|||:::||||||  :::::::::         ::
 945 ThrLysThrGlyAspIleLysGlyGlyIleGluSerThrSerGlyAsnVa 961

2655 ACACTTAAAAGACAGCGAATGGACGCTGCCGTCAGGCACGGAATTAGGCA 2704
     :::::::       |||      ||||||             :::::::::|
 961 lAsnIleThrAlaSerGlyAsnThrLeu...........LysValSerA 974

2705 ATTTAAACCTTGACAACGCCACCATTACACTCAATTCCGCCTATCGCCAC 2754
     ||:::::: :::::::::|||:::|||  :::
 974 snIleThrGlyGlnAspValThrValThrAlaAsp............... 985

2755 GATGCGGCAGGGGCGCAAACCGGCAGTGCGACAGATGCGCCGCGCCGCCG 2804
           |||||||||  |||  :::|||
 986 ......AlaGlyAlaLeuThrThrThrAla.................... 993

2805 TTCGCGCCGTTCGCGCCGTTCCCTATTATCCGTTACACCGCCAACTTCGG 2854
                       |||   :::|||:::|||    ::::::
 994 ................GlySerThrIleSerAlaThrThrGlyAsnAla. 1004

2855 TAGAATCCCGTTTCAACACGCTGACGGTAAACGGCAAATTGAACGGTCAG 2904
                     |||    |||  :::|||  :::||||||:::
1005 ..............AsnIleThrThrLysThrGlyAspIleAsnGlyLys 1016

2905 GGAACATTCCGCTTTATGTCGGAACTCTTCGGCTACCGCAGCGACAAATT 2954

1016 .................................................. 1016

2955 GAAGCTGGCGGAAAGTTCCGAAGGCACTTACACCTTGGCGGTCAACAAT. 3003
             :::|||||||||:::|||:::  ||||||:::::::::::::
1017 .......ValGluSerSerSerGlySerValThrLeuValAlaThrGlyA 1031

3004 ...........ACCGGCAACGAACCTGCAAGCCTCGAACAATTGACGGTA 3042
                :::||||||      ::::::         :::|||:::
1031 laThrLeuAlaValGlyAsnIleSerGlyAsnThrValThrIleThrAla 1047

3043 GTGGAAGGAAAAGACAACAAACCGCTGTCCGAAAACCTTAAT........ 3084
        :::||||||   ::::::  :::::::::::::::|||
1048 AspSerGlyLysLeuThrSerThrValGlySerThrIleAsnGlyThrAs 1064

3085 ..................................TTCACCCTGCAAA 3097
                                        |||::::::
1064 nSerValThrThrSerSerGlnSerGlyAspIleGluGlyThrIleSerG 1081

3098 ACGAACACGTCGATGCCGGCGCGTGGCGTTACCAACTCATCCGCAAAGAC 3147
     :::::  |||::::::  |||                        :::
1081 lyAsnThrValAsnValThrAla....................SerThr 1090

3148 GGCGAGTTCCGCCTGCATAATCCGGTCAAAGAACAAGAGCTTTCCGACAA 3197
     |||::::::  :::  |||  :::|||          ::::::  ||
1091 GlyAspLeuThrIleGlyAsnSerAlaLys.........ValGluAlaLy 1104

3198 ACTCGGCAAGGCAGAAGCCAAAAAACAGGCGGAAAAAGACAACGCGCAAA 3247
      |   |||  |||              ::::::  |||  ::::::|||:
1104 sAsnGlyAlaAlaThrLeuThrAlaGluSerGlyLysLeuThrThrGlnT 1121

3248 GCCTTGACGCGCTGATTGCGGCCGGGCGCGATGCCGTCGAAAAGACAGAA 3297
      ::   :::::::::  :::::::::               :::|||
1121 hrGlySerSerIleThrSerSerAsn............GlyGlnThrThr 1133

3298 AGCGTTGCCGAACCGGCCCGGCAGGCAGGCGGGGAAAAT........... 3336
        :::|||:::   :::      ||||||:::   |||
1134 LeuThrAlaLysAspSerSerIleAlaGlyAsnIleAsnAlaAlaAsnVa 1150

3337 .............GTCGGCATTATGCAGGCGGAGGAAGAGAAAAAACGGG 3373
                :::|||   :::   :::      :::::|||   :
1150 lThrLeuAsnThrThrGlyThrLeuThrThrThrGlyAspSerLys...I 1166

3374 TGCAGGCGGATAAAGACACCGCCTTGGCGAAACAGCGCGAAGCGGAAACC 3423
     :::::|||   :::   |||          :::   ::::::|||:::
1166 leAsnAlaThrSerGlyThrLeuThrIleAsnAlaLysAspAlaLysLeu 1182

3424 CGGCCGGCTACCACCGCCTTCCCCCGCGCCCGCCGCGCCCGCCGGGATTT 3473
           |||:::::::::
1183 AspGlyAlaAlaSerGly................................ 1188

3474 GCCGCAACTGCAACCCCAACCGCAGCCCCAACCGCAGCGCGACCTGATCA 3523
                                         :::|||   ::::::
1189 .................................AspArgThrValValA 1194

3524 GCCGTTATGCCAATAGCGGTTTGAGTGAATTTTCCGCCACGCTCAACAGC 3573
     ::          ||||||   ::::::   :::|||        :::|||
1194 snAlaThrAsnAlaSerGlySerGlyAsnValThrAlaLysThrSerSer 1210

3574 GTTTTCGCCGTACAGGACGAATTAGAC..............CGCGTATT 3608
           :::     :::|||:::            :::::::
1211 SerValAsnIleThrGlyAspLeuAsnThrIleAsnGlyLeuAsnIleIl 1227

3609 TGCCGAAGACCGCCGCAACGCCGTTTGGACAAGCGGCATCCGG...GACA 3655
     ::::|||:::   ||||||:::|||         |||   :::   |||:
1227 eSerGluAsnGlyArgAsnThrValArgLeuArgGlyLysGluIleAspV 1244

3656 CCAAACACTACCGTTCGCAAGATTTCCGCGCCTACCGCCAACAAACCGAC 3705
     ::|||:::           |||     ::::::   :::::::::::
1244 alLysTyrIle......GlnPro.GlyValAlaSerValGluGluValIl 1258

3706 CTGCGCCAAATCGGTATGCAGAAAAACC...................... 3733
           ||||||  |||   :::|||:::
1258 eGluAlaLysArgValLeuGluLysValLysAspLeuSerAspGluGluA 1275

3734 .....TCGGCAGCGGGCGCGTCGGCATCCTGTTTTCGCACAACCGGACCG 3778
           :::   |||      ::::::|||:::   |||   :::   |||
1275 rgGluThrLeuAlaLysLeuGlyValSerAlaValArgPheValGluPro 1291

3779 AAAACACCTTCGACGACGGCATCGGCAACTCGGCACGGCTTGCCCACGGC 3828
     ::::::       |||:::        |||::::::   :::   |||
1292 AsnAsnAlaIleThrVal......AsnThrGlnAsnGluPheThrThrLy 1306

3829 GCCGTTTTCGGGCAATACGGCATCGACAGGTTCTACA......TCGGCAT 3872
      |||   |||       |||   |||   |||::::::      |||::::
1306 sProSerSerGlnValThrIleSerGluGlyLysAlaCysPheSerSerG 1323

3873 CAGCGCGGGCGCGGGTTTTAGCAGCGGCAGCC 3904
      ::        |||||||||   :::   :::|||
1323 lyAsnGlyAlaArgValCysThrAsnValAla 1333
```

```
seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01827

seq_documentation_block:
ID   AAB01827 standard; Protein; 969 AA.
XX
AC   AAB01827;
XX
DT   11-SEP-2000  (first entry)
XX
DE   Haemophilus influenzae strain Joyc mature HMW2A protein, SEQ ID NO:32.
XX
KW   Mature HMW protein; hmw gene; hmwA1; hmwA2; high molecular weight;
KW   non-typeable Haemophilus influenzae; NTHi; non-encapsulated;
KW   recombinant production; Escherichia coli; antibacterial; vaccine;
KW   human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;
KW   detection; diagnosis.
XX
OS   Haemophilus influenzae strain Joyc.
XX
PN   WO200020609-A2.
XX
PD   13-APR-2000.
XX
PF   07-OCT-1999;   99WO-CA00938.
XX
PR   07-OCT-1998;   98US-0167568.
PR   08-DEC-1998;   98US-0206942.
XX
PA   (CONN-) CONNAUGHT LAB LTD.
XX
PI   Loosmore SM,  Yang Y,  Klein MH;
XX
DR   WPI; 2000-303789/26.
DR   N-PSDB; AAA52178.
XX
PT   Nucleic acid molecule for producing recombinant high molecular weight
PT   proteins of Haemophilus which are used as a vaccine to provide
PT   protection against Haemophilus induced diseases in humans -
XX
PS   Claim 8; Fig 19A-O; 307pp; English.
XX
CC   The invention relates to the recombinant production of Haemophilus
CC   influenzae high molecular weight (HMW) proteins in Escherichia coli. The
CC   expression construct used to effect recombinant expression comprises a
CC   promoter functional in E. coli (e.g., the T7 promoter) operably linked
CC   to a modified hmwABC operon from a non-typeable (non-encapsulated) H.
CC   influenzae (NTHi). Most HMW-expressing NTHi strains contain two hmw gene
CC   clusters termed hmw1ABC and hmw2ABC. Each hmwABC operon comprises hmwA,
CC   hmwB and hmwC genes. The hmwA genes encode the structural HMWA proteins
CC   and the hmwB and hmwC genes encode accessory proteins which are
CC   responsible for post-translational processing and secretion of the HMWA
CC   proteins. The modified hmwABC operon used in the expression construct of
CC   the invention contains an A gene modified such that it encodes only the
CC   mature HMWA. The invention also discloses hmwA genes (AAA52175-A52198)
CC   and HMWA proteins (AAB01824-B01849) from the non-typeable H. influenzae
CC   strains Joyc, K1, K21, LCDC2, PMH1, 15 and 12. The nucleic acids and
CC   vectors are used for the production of recombinant H. influenzae HMW
CC   proteins which can be used as vaccines to mediate a humoral or
CC   cell-mediated immune response to provide protection against diseases in
CC   humans caused by H. influenzae (e.g., otitis media, epiglottitis,
CC   pneumonia and tracheobronchitis). The HMW proteins are also useful as
CC   antigens in immunoassays for detecting antibodies against Haemophilus,
CC   HMW proteins and/or HMW peptides. The nucleotide sequences encoding the
CC   HMW proteins can be used to isolate and clone hmw genes from other
CC   non-typeable strains of Haemophilus via hybridisation reactions. The
CC   present sequence represents a mature HMWA protein from a non-typeable
CC   strain of H. influenzae.
XX
SQ   Sequence   969 AA;

alignment_scores:
           Quality:  280.50          Length:   1143
```

```
               Ratio:  0.506               Gaps:    52
 Percent Similarity:  48.469   Percent Identity: 20.385

alignment_block:
 US-09-303-518D-649 x AAB01827   ..

 Align seg 1/1  to: AAB01827  from: 1  to: 969

  1090 AATGTTTCTTTATCC..................GAGACAGCAAGAGAACC 1121
       |||||||||::::::                  :::   :::::|||
     3 AsnValSerIleGluAsnProSerThrGluArgAsnAspSerAsnGluAs 19

  1122 TGTTTATCATGCTGCAGGTGGTGTCAACAGTTATCGACCCAGACTGAATA 1171
        :::   :::::::::   |||   |||      :::|||:::::||||
    19 pLeuGluTyrThrGlyThrGlyGluAsnIleAsnAsnProLysValAsnA 36

  1172 ATGGA............................GAAAATATT...... 1185
       ||                               |||||||||
    36 snGlnSerLysLysThrLeuThrSerSerIleLeuGluAsnIleLeuLys 52

  1186 ......TCCTTTATTGAC...GAAGGAAAAGGCGAATTGATACTTACCAG 1226
             ||||||:::::      :::        ::::::   ::::::||
    53 LysGlySerPheValAsnIleThrAlaThrAspAsnIleTyrValAsnSe 69

  1227 CAACATCAATCAAGGT...GCTGGAGGATTATATTTCCAAGGAGATTTTA 1273
       |:::||||||   |||    :::|||   |||    ::::::|||
    69 rSerIleAsnIleGlyAspSerGlyHisLeuIleLeuSerGlyGlyGlyA 86

  1274 CGGTCTCGCCTGAAAATAACGAAACTTGGCAAGGCGCGGGCGTTCATATC 1323
                                   :::|||:::||||||   |||
    86 rg..........................AsnGlyGlyGlyValLysIle 93

  1324 AGTGAAGACAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAAACGACCG 1373
       :::::::::   |||   |||
    94 AsnLysAsnIleThrSerThr............................. 100

  1374 CCTGTCCAAAATCGGCAAAGGCACGCTGCACGTTCAAGCCAAAGGG.... 1419
                        |||:::|||   ::::::::::||||||
   101 ................GlyGlySerLeuThrIleAsnSerLysGlyTrpV 112

  1420 ..GAAAACCAAGGCTCGATCAGCGTGGGC..................... 1446
          :::   ::::::::||||||:::|||
   112 alAspIleHisSerAsnIleSerLeuGlyThrGlyPheLeuAsnIleThr 128

  1447 ...GACGGTACAGTCATTTTGGATCAGCAGGCAGACGATAAAGGCAAAAA 1493
          :::|||:::|||   ::::::::            ||||||:::::::
   129 SerAsnGlySerValAlaPheGluLysAlaAspLysAspLysAlaArgSe 145

  1494 ACAAGCCTTTAGTGAAATCGGCTTGGTCAGCGGCAGGGGTACGGTGCAAC 1543
       :   |||   ::::::|||                 ::::::|||   ::::::|
   145 rAlaAlaAspAlaGlnIle.........ValAlaGlnGlyIleIleAsnL 159

  1544 TGAATGCCGATAATCAGTTCAACCCCGACAAACTCTATTTCGGCTTTCGC 1593
       ||:::::::::|||:::                          ||||||
   159 euThrGlyGluAsnLys.......................ThrPheArg 167

  1594 GGCGGACGTTTGGATTTAAACGGGCATTCGCTTTCGTTCCACCGTATTCA 1643
          ::::::::::::|||||||||   :::   ::::::           ::
   168 LeuAsnAsnValSerLeuAsnGlyValGlyGlnGlyLeuSerIleThrSe 184

  1644 AAATACCGATGAAGGGGCGATGATTGTCAACCACAATCAAGACAAAGAAT 1693
       :|||                  :::   |||:::::::::      ::::
   184 rAsn..................ValGlyAsnGlnThrHisLysPheAspG 195

  1694 CCACCGTTACCATTACAGGCAATAAAGATATT..............GCT 1728
       ::   ::::::||||||||||||         |||              |||
   195 lyGluIleAsnIleThrGlyAsnValThrIleAsnGlnThrAlaProAla 211

  1729 ACAACCGGCAAT...AACAACAGCTTGGAT...............AGCAA 1760
       ||||||:::      |||   |||   |||                |||
   212 ThrThrAlaTyrTrpAsnPheSerTyrAspSerTyrTrpAsnValSerTh 228

  1761 AAAAGAAATTGCCTACAACGGTTGGTTTGGCGAGAAAGATACGACCAAAA 1810
```

```
           :::::      |||:::   |||             |||::::
 228 rLeuAsnValGlnLysAsnSerSerPheThrPheIleLysArgThrGluS 245

1811 CGAACGGGCGGCTCAACCTTGTTTACCAGCCCGCCGCAGAAGACCGCACC 1860
     ::|||                     :::   |||:::::      |||:::
 245 erAsnArg...............PheGlyProThrThrProLeuArgSer 256

1861 CTGCTGCTTTCCGGCGGA.........ACAAATTTAAACGGCAACATC.. 1899
              |||||||||          ::::::   |||||||||:::
 257 .........SerGlyGlyValPhePheAsnGlyThrAsnGlyAsnMetVa 270

1900 .......ACGCAAACAAACGGCAAACTGTTTTTCAGCGGCAGACCAACAC 1942
           :::   |||||::::::::::::|||:::    :::|||
 270 lLeuAsnValGlyThrAsnSerArgValLeuPheAsnLeuLysPro.... 285

1943 CGCACGCCTACAATCATTTAAACGACCATTGGTCGCAAAAAGAGGGCATT 1992
                                |||::::::   :::::::::::   :::
 286 ....................AsnGluAsnThrAsnAsnSerLysProLeu 295

1993 CCTCGCGGGGAAATCGTGTGGGACAACGACTGGATCAACCGCACATTTAA 2042
     |||                                         |||::
 296 Pro..................................LeuGlnPheAs 300

2043 AGCGGAAAACTTCCAAATTAAAGGCGGACAGGCGGTGGTTTCCCGCAATG 2092
     :|||:::                 |||   |||||
 300 nAlaAsnIleThrAlaIleGlyGlyGly..................... 309

2093 TTGCCAAAGTGAAAGGCGATTGGCATTTGAGCAATCACGCCCAAGCAGTT 2142
           :::|||:::   |||   |||    :::|||||||:::
 310 .....SerValSerPheAspIleHis...AlaAsnHisSer......... 320

2143 TTTGGTGTCGCACCGCATCAAAGCCACACAATCTGTACACGTTCGGACTG 2192
       |||    :::   :::   :::   :::        :::   ||||||
 321 ...GlyArgGlyAlaGluLeuLysMetAsnThrIleAsnIleSerAspGl 336

2193 GACGGGTCTGACA............................AATTGTG 2212
      |||:::||||||                                 :
 336 yThrSerLeuThrLeuGlnSerHisValArgLysAspSerAlaPheIleI 353

2213 TCGAAAAAACCATTACC................................ 2229
     ::::::|||   :::|||
 353 leSerLysAspLeuThrIleAsnAlaThrGlySerAsnPheThrLeuGlu 369

2230 .............................GACGATAAAGTGATTGCTTC 2249
                                      ::::::|||:::||
 370 GlnSerProAspSerPheThrAspLysTyrProGlyArgAlaIleSerSe 386

2250 ATTGACTAAGACCGACATCAGC...GGCAATGTCGATCTTGCCGATCACG 2296
     |      :::        ||||||   |||||||||:::|||:::    :::
 386 rThrLysAsnIleThrIleSerGlyGlyAsnValSerLeuGlyGlyGln. 402

2297 CTCATTTAAATCTCACAGGGCTTGCCACACTCAACGGCAATCTTAGTGCA 2346
                      ::::::   :::   ::::::||||||::::::
 403 .............AsnSerSerSerAspIleLysGlyAsnIleThrIle 414

2347 AATGGCGATACACGTTATACAGTC...AGCCACAACGCCACCCAAAAC.. 2391
     :::::::::|||:::   |||:::   :::||||||:::   ::::::
 415 LysSerSerThrAsnValThrLeuLysAlaHisAsnSerProArgAspPh 431

2392 ......................GGCAACCTTAGCCTCGTGGGCAATGCCC 2419
                          |||||||||::::::   ||||||
 431 eAlaSerArgThrLeuThrLeuGlyAsnLeuAsnValGluGlyAsnLeuT 448

2420 AAGCAACATTTAATCAAGCCACATTAAACGGCAACACATCGGCTTCGGGC 2469
        |||   :::   |||   ::::::||||||   |||      :::
 448 hrLeuThrGlySerValAlaAspIleLysGlyAsnLeuSerIleLeuAsn 464

2470 AATGCTTCATTT......AATCTAAGCGACCACGCCGTACAAAACGGCAG 2513
     :::|||:::|||       :::   |||::::::      :::|||::
 465 AspAlaThrPheLysGlyGluThrSerGluAsnLeuAsnIleThrGlyAs 481

2514 TCTGACGCTTTCCGGCAACGCTAAGGCAAACGTAAGCCATTCCGCACTCA 2563
     ::::|||   :::|||:::|||         |||::::::::::::::::::|

 481 nPheThrAsnAsnGlyThrAlaAspIleAsnIleLysGlnGlyValValA 498

2564 ACGGTAATGTCTCCCTAGCCGATAAGGCAGTATTCCATTTTGAAAGCAGC 2613
     ||    :::   ::::::::::::|||:::    :::::::::    ::::::
 498 snIleGlnGlyAsnIleThrAsnLysGlyGlyLeuAsnIleThrThrAsn 514

2614 CGCTTTACCGGACAA.........ATCAGCGGC..............GG 2639
         ::::::|||          |||:::|||                 ||
 515 AlaGlnAsnAsnGlnLysThrIleIleAsnGlyAsnIleThrAsnGluGl 531

2640 CAAGGATACGGCATTACACTTAAAAGACAGCGAATGGACGCTGCCGTCAG 2689
     |     |||        |||::::::|||||||||:::              ::::
 531 yGlyAsp......LeuAsnIleLysAspSerAsn............AsnA 542

2690 GCACGGAATTA.........GGCAATTTAAACCTTGACAACGCCACCATT 2730
     ::::::|||:::         ||||||::::::      ::::::::::::
 542 snAlaGluIleGlnIleGlyGlyAsnIleSerGlnLysLysGlyAsnLeu 558

2731 ACACTCAATTCCGCCTATCGCCACGAT....................... 2757
     |||::::::|||             |||
 559 ThrIleSerSer............AspLysIleAsnIleThrLysLysIl 571

2758 .......GCGGCAGGGGCGCAAACCGGCAGTGCGACAGATGCGCCGCGCC 2800
                ||||||::::::    |||::::::    ::::::|||
 571 eThrIleLysAlaGlyValAspGluGlyGlySerAspSerSerProAlaS 588

2801 GCCGTTCGCGCCGTTCGCGCCGTTCCCTATTATCCGTTACACCGCCAACT 2850
       ::::::
 588 erAsnAla......................................... 590

2851 TCGGTAGAATCCCGTTTCAACACGCTGACGGTAAACGGCAAATTGAACGG 2900
     ::::::          :::   :::||||||   ::::::|||   ||||||
 591 AsnLeuThrIleLysThrLysThrLeuGluLeuThrGlyAspLeuAsnIl 607

2901 TCAGGGAACATTCCGCTTTATGTCGGAACTCTTCGGCTACCGCAGCGACA 2950
       :::|||        :::       :::|||:::   :::    :::::::::
 607 eSerGlyPheAsnLys......AlaGluIleThrAlaLysAsnGlyAsnA 622

2951 AATTGAAGCTGGCGGAAAGTTCCGAAGGCACTTACACCTTGGCGGTCAAC 3000
       |||    :::::::::::|||:::|||:::
 622 spLeuThrIleGlyLysAlaSerAspGlyAsn.................. 632

3001 AATACCGGCAACGAACCTGCAAGCCTCGAACAATTGACGGTAGTGGAAGG 3050
                          |||:::    :::::::::|||      :::
 633 ..................AlaAsnAlaLysLysValThrPheAspLysVa 643

3051 AAAAGACAACAAACCGCTGTCCGAAAACCTTAATTTCACCCTGCAAAACG 3100
       ||||||:::|||      :::::::::    |||   |||
 643 lLysAspSerLysIleSerAlaAsnGlyHisAsnValThr.......... 656

3101 AACACGTCGATGCCGGCGCGTGGCGTTACCAACTCATCCGCAAAGACGGC 3150

 656 ................................................. 656

3151 GAGTTCCGCCTGCATAATCCGGTCAAAGAACAAGAGCTTTCCGACAAACT 3200
                                      |||::::::|||::
 657 ..................................LeuAsnSerLysVa 661

3201 CGGCAAGGCAGAAGCCAAAAAACAGGCGGAAAAAGACAACGCGCAAAGCC 3250
     :      :::::::::   ::::::|||:::    :::|||    ::::::
 661 lGluThrSerAsnSerAspSerSerAlaAspAspSerAsnAspAsnAsnT 678

3251 TTGACGCGCTGATTGCGGCCGGGCGCGATGCCGTCGAAAAGACAGAAAGC 3300
                |||:::|||
 678 hrGlyLeuThrIleSerAla............................ 684

3301 GTTGCCGAACCGGCCCGGCAGGCAGGCGGGGAAAATGTCGGCATTATGCA 3350
                                  ::::::|||    :::     ::
 685 ............................LysAspValThrValAsnAs 691

3351 GGCGGAGGAAGAGAAAAAACGGGTGCAGGCGGATAAAGACACCGCCTTGG 3400
     :          :::   |||    ::::::   :::       |||:::    :
 691 nAspValThrSerHisLysThrIleAsnIleSerAlaThrThrGlyAsnV 708
```

```
3401 CGAAACAGCGCGAAGCGGAAACCCGGCCGGCTACCACCGCCTTCCCCCGC 3450
      ::      :::|||:::    |||      |||:::|||:::  |||
 708 alThrThrLysGluSerThrThrIleAsnAlaAlaThrGly.SerValGl 724

3451 GCCCGCCGCGCCCGCCGGGATTTGCCGCAACTGCAACCCCAACCGCAGCC 3500
        :::|||    :::|||          :::          :::::::::
 724 uValThrAlaLysThrGlyAspIleSerGlyThrIleSerGlyAsnThrV 741

3501 CCAACCGCAGCGCGACCTGATCAGCCGTTATGCCAATAGCGGTTTGAGTG 3550
       |||   :::||||||   |||   :::         :::
 741 alAsnValThrAlaThrAspSerLeuThrThrGlnAlaSer......... 754

3551 AATTTTCCGCCACGCTCAACAGCGTTTTCGCCGTACAGGACGAATTAGAC 3600
                 |||   :::   |||       :::|||:::   ||
 755 ...........SerSerIleThrSerSerAsnGlyGlnThrThrLeuTh 767

3601 CGCGTATTTGCCGAAGACCGCCGCAACGCCGTTTGGACAAGCGGCATCCG 3650
      ||||           :::    |||:::::         :::||||||:::
 767 rAlaLysAsnGlySerIleAlaGlySerIle......AspAlaAlaAsnV 782

3651 GGACACCAAACACT......ACCGTTCGCAAGATTTCCGCGCCTACCGCC 3694
        |||   ||||||       |||:::       :::::::::    :::
 782 alThrLeuAsnThrThrGlyThrLeuThrThrValAlaGlySerAsnIle 798

3695 AACAAACCGACCTGCGCCAAATCGGTATGCAGAAAAACCTCGGCAGCGGG 3744
      :::     :::    :::    ::::::        |||    :::
 799 LysAlaThrSerGlyThrLeuAlaIleAsnAlaLysAspAlaLysLeuAs 815

3745 CGCGTCGGCATCCTGTTTTCGCACAACCGGACCGAAAACACCTTCGACGA 3794
       ::::::||||||                |||    :::|||       ::::
 815 pGlyThrAlaSer...............GlyAspArgThrValValAsnA 827

3795 CGGCATCGGCAACTCGGCACGGCTTGCCCACGGCGCCGTTTTCGGGCAAT 3844
      :::::::|||:::        |||        ||||||      |||:::|||
 827 laThrAsnAlaSerGlySerGlySerValThrAlaAlaThrSerSerAsn 843

3845 ACGGCATCGACAGGTTCTACATCGGCATCAGCGCGGGCGCGGGTTTTAGC 3894
      :::       ||||||:::   |||:::               :::::::::::
 844 ValAsnIleThrGlyAspLeuSerThrIleAsnGlyLeuAsnIleIleSe 860

3895 AGCGGCAGCCTTTCAGACGGCATCGGAGGCAAAATCCGCCGCCGCGTGCT 3944
      :       :::    :::|||:::      :::    ::::::|||
 860 rLysAsnGlyLysAsnThrVal......ValLeuLysGlyAlaGluIleA 875

3945 GCATTACGGCATTCAGGCACGATACCGCGCCGGTTTCGGCGGATTCGGCA 3994
       :::       ::::::      :::|||                ||||||
 875 spValLysTyrIleGlnProGlyValAla..............SerAla 886

3995 TCGAACCGCACATCGGCGCAACGCGCT.................ATTTC 4026
      ::::::       :::|||:::||||||                 :::||
 887 AsnGluValIleGluAlaLysArgAlaLeuGluLysValLysAspLeuSe 903

4027 GTCCAAAAAGCGGATTACCGCTACGAAAACGTCAATATCGCCACCCCCGG 4076
      |:::::::::|||    |||     :::|||    :::::::|||
 903 rAspGluGluArgGluThrLeuAlaLysLeuGlyValSerAlaValArgP 920

4077 CCTTGCATTCAACCGCTACCGCGCGGGCATTAAGGCAGATTATTCATTCA 4126
        ::::::    :::    |||                          |||
 920 heValGluProAsnAsnThr....................IleThrVal 929

4127 AACCGGCGCAACACATTTCCATCACGC 4153
      |||    :::|||    |||    :::|||
 930 AsnThrGlnAsnGluPheThrThrArg 938
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01826

```
seq_documentation_block:
ID   AAB01826 standard; Protein; 975 AA.
XX
AC   AAB01826;
XX
DT   11-SEP-2000  (first entry)
XX
DE   Haemophilus influenzae strain Joyc HMW2A protein, SEQ ID NO:30.
XX
KW   HMW protein; hmw gene; hmwA1; hmwA2; high molecular weight;
KW   non-typeable Haemophilus influenzae; NTHi; non-encapsulated;
KW   recombinant production; Escherichia coli; antibacterial; vaccine;
KW   human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;
KW   detection; diagnosis.
XX
OS   Haemophilus influenzae strain Joyc.
XX
PN   WO200020609-A2.
XX
PD   13-APR-2000.
XX
PF   07-OCT-1999;   99WO-CA00938.
XX
PR   07-OCT-1998;   98US-0167568.
PR   08-DEC-1998;   98US-0206942.
XX
PA   (CONN-) CONNAUGHT LAB LTD.
XX
PI   Loosmore SM,  Yang Y,  Klein MH;
XX
DR   WPI; 2000-303789/26.
DR   N-PSDB; AAA52177.
XX
PT   Nucleic acid molecule for producing recombinant high molecular weight
PT   proteins of Haemophilus which are used as a vaccine to provide
PT   protection against Haemophilus induced diseases in humans -
XX
PS   Claim 12; Fig 19A-O; 307pp; English.
XX
CC   The invention relates to the recombinant production of Haemophilus
CC   influenzae high molecular weight (HMW) proteins in Escherichia coli. The
CC   expression construct used to effect recombinant expression comprises a
CC   promoter functional in E. coli (e.g., the T7 promoter) operably linked
CC   to a modified hmwABC operon from a non-typeable (non-encapsulated) H.
CC   influenzae (NTHi). Most HMW-expressing NTHi strains contain two hmw gene
CC   clusters termed hmw1ABC and hmw2ABC. Each hmwABC operon comprises hmwA,
CC   hmwB and hmwC genes. The hmwA genes encode the structural HMWA proteins
CC   and the hmwB and hmwC genes encode accessory proteins which are
CC   responsible for post-translational processing and secretion of the HMWA
CC   proteins. The modified hmwABC operon used in the expression construct of
CC   the invention contains an A gene modified such that it encodes only the
CC   mature HMWA. The invention also discloses hmwA genes (AAA52175-A52198)
CC   and HMWA proteins (AAB01824-B01849) from the non-typeable H. influenzae
CC   strains Joyc, K1, K21, LCDC2, PMH1, 15 and 12. The nucleic acids and
CC   vectors are used for the production of recombinant H. influenzae HMW
CC   proteins which can be used as vaccines to mediate a humoral or
CC   cell-mediated immune response to provide protection against diseases in
CC   humans caused by H. influenzae (e.g., otitis media, epiglottitis,
CC   pneumonia and tracheobronchitis). The HMW proteins are also useful as
CC   antigens in immunoassays for detecting antibodies against Haemophilus,
CC   HMW proteins and/or HMW peptides. The nucleotide sequences encoding the
CC   HMW proteins can be used to isolate and clone hmw genes from other
CC   non-typeable strains of Haemophilus via hybridisation reactions. The
CC   present sequence represents an HMWA protein from a non-typeable strain of
CC   H. influenzae.
XX
SQ   Sequence   975 AA;

alignment_scores:
            Quality:  280.50             Length:   1143
              Ratio:  0.506                Gaps:     52
 Percent Similarity:  48.469   Percent Identity: 20.385

alignment_block:
 US-09-303-518D-649 x AAB01826   ..

 Align seg 1/1  to: AAB01826  from: 1  to: 975
```

```
1090 AATGTTTCTTTATCC..................GAGACAGCAAGAGAACC 1121
     |||||||||:::::                    :::   ::::::|||
   9 AsnValSerIleGluAsnProSerThrGluArgAsnAspSerAsnGluAs 25

1122 TGTTTATCATGCTGCAGGTGGTGTCAACAGTTATCGACCCAGACTGAATA 1171
      :::   :::::::::   |||   |||      :::|||::::::||||
  25 pLeuGluTyrThrGlyThrGlyGluAsnIleAsnAsnProLysValAsnA 42

1172 ATGGA.............................GAAAATATT...... 1185
     ||                                 ||||||||
  42 snGlnSerLysLysThrLeuThrSerSerIleLeuGluAsnIleLeuLys 58

1186 ......TCCTTTATTGAC...GAAGGAAAAGGCGAATTGATACTTACCAG 1226
           ||||||::::::      :::      ::::::   :::::::||
  59 LysGlySerPheValAsnIleThrAlaThrAspAsnIleTyrValAsnSe 75

1227 CAACATCAATCAAGGT...GCTGGAGGATTATATTTCCAAGGAGATTTTA 1273
     |:::|||||||   |||   :::|||    |||    :::::::|||
  75 rSerIleAsnIleGlyAspSerGlyHisLeuIleLeuSerGlyGlyGlyA 92

1274 CGGTCTCGCCTGAAAATAACGAAACTTGGCAAGGCGCGGGCGTTCATATC 1323
                                  :::|||:::||||||   |||
  92 rg..........................AsnGlyGlyGlyValLysIle 99

1324 AGTGAAGACAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAAACGACCG 1373
     :::::::::   |||   |||
 100 AsnLysAsnIleThrSerThr............................ 106

1374 CCTGTCCAAAATCGGCAAAGGCACGCTGCACGTTCAAGCCAAAGGG.... 1419
                       |||:::|||   :::::::::||||||
 107 ................GlyGlySerLeuThrIleAsnSerLysGlyTrpV 118

1420 ..GAAAACCAAGGCTCGATCAGCGTGGGC..................... 1446
       :::    :::::::::||||||:::|||
 118 alAspIleHisSerAsnIleSerLeuGlyThrGlyPheLeuAsnIleThr 134

1447 ...GACGGTACAGTCATTTTGGATCAGCAGGCAGACGATAAAGGCAAAAA 1493
        :::|||:::|||   :::::::::            ||||||:::::::
 135 SerAsnGlySerValAlaPheGluLysAlaAspLysAspLysAlaArgSe 151

1494 ACAAGCCTTTAGTGAAATCGGCTTGGTCAGCGGCAGGGGTACGGTGCAAC 1543
     :    |||    ::::::|||                :::::::|||    ::::::|
 151 rAlaAlaAspAlaGlnIle.........ValAlaGlnGlyIleIleAsnL 165

1544 TGAATGCCGATAATCAGTTCAACCCCGACAAACTCTATTTCGGCTTTCGC 1593
     ||:::::::::|||:::                        ||||||
 165 euThrGlyGluAsnLys.......................ThrPheArg 173

1594 GGCGGACGTTTGGATTTAAACGGGCATTCGCTTTCGTTCCACCGTATTCA 1643
         ::::::::::::|||||||||   :::     ::::::      ::
 174 LeuAsnAsnValSerLeuAsnGlyValGlyGlnGlyLeuSerIleThrSe 190

1644 AAATACCGATGAAGGGGCGATGATTGTCAACCACAATCAAGACAAAGAAT 1693
     :|||                  :::    |||:::::::::       ::::
 190 rAsn.................ValGlyAsnGlnThrHisLysPheAspG 201

1694 CCACCGTTACCATTACAGGCAATAAAGATATT..............GCT 1728
     ::    ::::::||||||||||||         |||            |||
 201 lyGluIleAsnIleThrGlyAsnValThrIleAsnGlnThrAlaProAla 217

1729 ACAACCGGCAAT...AACAACAGCTTGGAT..............AGCAA 1760
     ||||||:::       |||    |||    |||              |||
 218 ThrThrAlaTyrTrpAsnPheSerTyrAspSerTyrTrpAsnValSerTh 234

1761 AAAAGAAATTGCCTACAACGGTTGGTTTGGCGAGAAAGATACGACCAAAA 1810
         ::::::        |||:::    |||              |||:::::
 234 rLeuAsnValGlnLysAsnSerSerPheThrPheIleLysArgThrGluS 251

1811 CGAACGGGCGGCTCAACCTTGTTTACCAGCCCGCCGCAGAAGACCGCACC 1860
     ::|||                 :::    |||::::::     |||:::
 251 erAsnArg..............PheGlyProThrThrProLeuArgSer 262

1861 CTGCTGCTTTCCGGCGGA.........ACAAATTTAAACGGCAACATC.. 1899

              |||||||||         ::::::   ||||||||||:::
 263 .........SerGlyGlyValPhePheAsnGlyThrAsnGlyAsnMetVa 276

1900 .......ACGCAAACAAACGGCAAACTGTTTTTCAGCGGCAGACCAACAC 1942
          :::    ||||||:::::::::::::|||:::    :::|||
 276 lLeuAsnValGlyThrAsnSerArgValLeuPheAsnLeuLysPro.... 291

1943 CGCACGCCTACAATCATTTAAACGACCATTGGTCGCAAAAAGAGGGCATT 1992
                                 |||:::::   :::::::::::   :::
 292 ...................AsnGluAsnThrAsnAsnSerLysProLeu 301

1993 CCTCGCGGGGAAATCGTGTGGGACAACGACTGGATCAACCGCACATTTAA 2042
     |||                                       |||::
 302 Pro...................................LeuGlnPheAs 306

2043 AGCGGAAAACTTCCAAATTAAAGGCGGACAGGCGGTGGTTTCCCGCAATG 2092
     :|||:::          |||    ||||||
 306 nAlaAsnIleThrAlaIleGlyGlyGly..................... 315

2093 TTGCCAAAGTGAAAGGCGATTGGCATTTGAGCAATCACGCCCAAGCAGTT 2142
          :::|||:::    |||    |||    :::|||||||:::
 316 .....SerValSerPheAspIleHis...AlaAsnHisSer......... 326

2143 TTTGGTGTCGCACCGCATCAAAGCCACACAATCTGTACACGTTCGGACTG 2192
        |||    :::    :::    :::    :::         :::    ||||||
 327 ...GlyArgGlyAlaGluLeuLysMetAsnThrIleAsnIleSerAspGl 342

2193 GACGGGTCTGACA.............................AATTGTG 2212
     |||:::||||||                                  :
 342 yThrSerLeuThrLeuGlnSerHisValArgLysAspSerAlaPheIleI 359

2213 TCGAAAAAAACCATTACC................................ 2229
     ::::::|||    :::|||
 359 leSerLysAspLeuThrIleAsnAlaThrGlySerAsnPheThrLeuGlu 375

2230 .............................GACGATAAAGTGATTGCTTC 2249
                                    ::::::|||:::||
 376 GlnSerProAspSerPheThrAspLysTyrProGlyArgAlaIleSerSe 392

2250 ATTGACTAAGACCGACATCAGC...GGCAATGTCGATCTTGCCGATCACG 2296
     |        :::          ||||||    |||||||||:::|||:::    :::
 392 rThrLysAsnIleThrIleSerGlyGlyAsnValSerLeuGlyGlyGln. 408

2297 CTCATTTAAATCTCACAGGGCTTGCCACACTCAACGGCAATCTTAGTGCA 2346
                  ::::::    :::    ::::::||||||:::::::
 409 ..............AsnSerSerSerAspIleLysGlyAsnIleThrIle 420

2347 AATGGCGATACACGTTATACAGTC...AGCCACAACGCCACCCAAAAC.. 2391
     :::::::::|||:::    |||:::    :::|||||||:::    ::::::
 421 LysSerSerThrAsnValThrLeuLysAlaHisAsnSerProArgAspPh 437

2392 ......................GGCAACCTTAGCCTCGTGGGCAATGCCC 2419
                            |||||||||::::::    ||||||
 437 eAlaSerArgThrLeuThrLeuGlyAsnLeuAsnValGluGlyAsnLeuT 454

2420 AAGCAACATTTAATCAAGCCACATTAAACGGCAACACATCGGCTTCGGGC 2469
        |||    :::    |||    ::::::||||||    |||         :::
 454 hrLeuThrGlySerValAlaAspIleLysGlyAsnLeuSerIleLeuAsn 470

2470 AATGCTTCATTT......AATCTAAGCGACCACGCCGTACAAAACGGCAG 2513
     :::|||:::|||        :::    |||::::::        :::|||::
 471 AspAlaThrPheLysGlyGluThrSerGluAsnLeuAsnIleThrGlyAs 487

2514 TCTGACGCTTTCCGGCAACGCTAAGGCAAACGTAAGCCATTCCGCACTCA 2563
     ::::|||    :::|||:::|||         |||::::::::::::::::::|
 487 nPheThrAsnAsnGlyThrAlaAspIleAsnIleLysGlnGlyValValA 504

2564 ACGGTAATGTCTCCCTAGCCGATAAGGCAGTATTCCATTTTGAAAGCAGC 2613
     ||    :::    ::::::::::::|||:::    :::::::::    ::::::
 504 snIleGlnGlyAsnIleThrAsnLysGlyGlyLeuAsnIleThrThrAsn 520

2614 CGCTTTACCGGACAA.........ATCAGCGGC...............GG 2639
          ::::::|||          |||:::|||                ||
```

```
 521 AlaGlnAsnAsnGlnLysThrIleIleAsnGlyAsnIleThrAsnGluGl 537

2640 CAAGGATACGGCATTACACTTAAAAGACAGCGAATGGACGCTGCCGTCAG 2689
     |  |||          |||::::::||||||||:::        ::::
 537 yGlyAsp......LeuAsnIleLysAspSerAsn...........AsnA 548

2690 GCACGGAATTA.........GGCAATTTAAACCTTGACAACGCCACCATT 2730
     :::::|||:::         ||||||:::::    :::::::::::
 548 snAlaGluIleGlnIleGlyGlyAsnIleSerGlnLysLysGlyAsnLeu 564

2731 ACACTCAATTCCGCCTATCGCCACGAT....................... 2757
     |||::::::|||           |||
 565 ThrIleSerSer...........AspLysIleAsnIleThrLysLysIl 577

2758 .......GCGGCAGGGGCGCAAACCGGCAGTGCGACAGATGCGCCGCGCC 2800
              ||||||:::::   |||:::::  ::::::|||
 577 eThrIleLysAlaGlyValAspGluGlyGlySerAspSerSerProAlaS 594

2801 GCCGTTCGCGCCGTTCGCGCCGTTCCCTATTATCCGTTACACCGCCAACT 2850
       ::::::
 594 erAsnAla.......................................... 596

2851 TCGGTAGAATCCCGTTTCAACACGCTGACGGTAAACGGCAAATTGAACGG 2900
     ::::::    :::   :::||||||  ::::::|||  ||||||
 597 AsnLeuThrIleLysThrLysThrLeuGluLeuThrGlyAspLeuAsnIl 613

2901 TCAGGGAACATTCCGCTTTATGTCGGAACTCTTCGGCTACCGCAGCGACA 2950
       :::|||    :::     :::|||:::   :::   :::::::::
 613 eSerGlyPheAsnLys......AlaGluIleThrAlaLysAsnGlyAsnA 628

2951 AATTGAAGCTGGCGGAAAGTTCCGAAGGCACTTACACCTTGGCGGTCAAC 3000
       |||   ::::::::::::|||:::|||:::
 628 spLeuThrIleGlyLysAlaSerAspGlyAsn.................. 638

3001 AATACCGGCAACGAACCTGCAAGCCTCGAACAATTGACGGTAGTGGAAGG 3050
                       |||:::  :::::::::|||      :::
 639 ..................AlaAsnAlaLysLysValThrPheAspLysVa 649

3051 AAAAGACAACAAACCGCTGTCCGAAAACCTTAATTTCACCCTGCAAAACG 3100
       ||||||:::|||      :::::::::   |||   |||
 649 lLysAspSerLysIleSerAlaAsnGlyHisAsnValThr.......... 662

3101 AACACGTCGATGCCGGCGCGTGGCGTTACCAACTCATCCGCAAAGACGGC 3150

 662 .................................................. 662

3151 GAGTTCCGCCTGCATAATCCGGTCAAAGAACAAGAGCTTTCCGACAAACT 3200
                                       |||::::::|||::
 663 ..................................LeuAsnSerLysVa 667

3201 CGGCAAGGCAGAAGCCAAAAAACAGGCGGAAAAAGACAACGCGCAAAGCC 3250
     :     :::::::::   ::::::|||:::   :::|||   ::::::
 667 lGluThrSerAsnSerAspSerSerAlaAspAspSerAsnAspAsnAsnT 684

3251 TTGACGCGCTGATTGCGGCCGGGCGCGATGCCGTCGAAAAGACAGAAAGC 3300
                 |||:::|||
 684 hrGlyLeuThrIleSerAla............................ 690

3301 GTTGCCGAACCGGCCCGGCAGGCAGGCGGGGAAAATGTCGGCATTATGCA 3350
                                  ::::::|||   :::   ::
 691 ............................LysAspValThrValAsnAs 697

3351 GGCGGAGGAAGAGAAAAAACGGGTGCAGGCGGATAAAGACACCGCCTTGG 3400
     :         :::   |||   ::::::   :::      |||:::   :
 697 nAspValThrSerHisLysThrIleAsnIleSerAlaThrThrGlyAsnV 714

3401 CGAAACAGCGCGAAGCGGAAACCCGGCCGGCTACCACCGCCTTCCCCCGC 3450
      ::      :::|||:::   |||      |||:::|||:::  |||
 714 alThrThrLysGluSerThrThrIleAsnAlaAlaThrGly.SerValGl 730

3451 GCCCGCCGCGCCCGCCGGGATTTGCCGCAACTGCAACCCCAACCGCAGCC 3500
          :::|||   :::|||         :::         :::::::::
 730 uValThrAlaLysThrGlyAspIleSerGlyThrIleSerGlyAsnThrV 747

3501 CCAACCGCAGCGCGACCTGATCAGCCGTTATGCCAATAGCGGTTTGAGTG 3550
       |||   :::||||||   |||   :::         :::
 747 alAsnValThrAlaThrAspSerLeuThrThrGlnAlaSer......... 760

3551 AATTTTCCGCCACGCTCAACAGCGTTTTCGCCGTACAGGACGAATTAGAC 3600
                     |||   :::   |||      :::|||:::   ||
 761 ...........SerSerIleThrSerSerAsnGlyGlnThrThrLeuTh 773

3601 CGCGTATTTGCCGAAGACCGCCGCAACGCCGTTTGGACAAGCGGCATCCG 3650
     ||||          :::   |||:::::         :::||||||:::
 773 rAlaLysAsnGlySerIleAlaGlySerIle......AspAlaAlaAsnV 788

3651 GGACACCAAACACT......ACCGTTCGCAAGATTTCCGCGCCTACCGCC 3694
      |||   ||||||         |||:::        :::::::::   :::
 788 alThrLeuAsnThrThrGlyThrLeuThrThrValAlaGlySerAsnIle 804

3695 AACAAACCGACCTGCGCCAAATCGGTATGCAGAAAAACCTCGGCAGCGGG 3744
     :::         :::   :::   ::::::         |||   :::
 805 LysAlaThrSerGlyThrLeuAlaIleAsnAlaLysAspAlaLysLeuAs 821

3745 CGCGTCGGCATCCTGTTTTCGCACAACCGGACCGAAAACACCTTCGACGA 3794
      ::::::||||||                  |||   :::|||      ::::
 821 pGlyThrAlaSer..............GlyAspArgThrValValAsnA 833

3795 CGGCATCGGCAACTCGGCACGGCTTGCCCACGGCGCCGTTTCGGGCAAT 3844
     :::::::|||:::         |||         ||||||    |||:::|||
 833 laThrAsnAlaSerGlySerGlySerValThrAlaAlaThrSerSerAsn 849

3845 ACGGCATCGACAGGTTCTACATCGGCATCAGCGCGGGCGCGGGTTTTAGC 3894
     :::         ||||||:::   |||:::            :::::::::::
 850 ValAsnIleThrGlyAspLeuSerThrIleAsnGlyLeuAsnIleIleSe 866

3895 AGCGGCAGCCTTTCAGACGGCATCGGAGGCAAAATCCGCCGCCGCGTGCT 3944
     :        :::   :::|||:::      :::   ::::::|||
 866 rLysAsnGlyLysAsnThrVal......ValLeuLysGlyAlaGluIleA 881

3945 GCATTACGGCATTCAGGCACGATACCGCGCCGGTTTCGGCGGATTCGGCA 3994
        :::      ::::::      :::|||             ||||||
 881 spValLysTyrIleGlnProGlyValAla..............SerAla 892

3995 TCGAACCGCACATCGGCGCAACGCGCT.................ATTTC 4026
     ::::::         :::|||:::||||||                 :::||
 893 AsnGluValIleGluAlaLysArgAlaLeuGluLysValLysAspLeuSe 909

4027 GTCCAAAAAGCGGATTACCGCTACGAAAACGTCAATATCGCCACCCCCGG 4076
     |:::::::::|||   |||   :::|||   ::::::|||
 909 rAspGluGluArgGluThrLeuAlaLysLeuGlyValSerAlaValArgP 926

4077 CCTTGCATTCAACCGCTACCGCGCGGGCATTAAGGCAGATTATTCATTCA 4126
        ::::::   :::   |||                       |||
 926 heValGluProAsnAsnThr....................IleThrVal 935

4127 AACCGGCGCAACACATTTCCATCACGC 4153
     |||   :::|||   |||   :::|||
 936 AsnThrGlnAsnGluPheThrThrArg 944
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:AAW30294

```
seq_documentation_block:
ID   AAW30294 standard; Protein; 1477 AA.
XX
AC   AAW30294;
XX
DT   14-APR-1998  (first entry)
XX
DE   Non-typeable Haemophilus high mol.wt. surface protein HMW2.
XX
KW   Non-typeable Haemophilus; high molecular weight surface protein;
KW   HMW2; hmw2A gene; immunogen; vaccine; otitis media.
XX
OS   Haemophilus influenzae strain 12.
XX
```

```
FH   Key                 Location/Qualifiers
FT   Misc-difference     34
FT                       /note= "encoded by TTC"
FT   Misc-difference     35
FT                       /note= "encoded by CGC"
FT   Misc-difference     36
FT                       /note= "encoded by TAT"
FT   Misc-difference     37
FT                       /note= "encoded by GTT"
FT   Misc-difference     38
FT                       /note= "encoded by ACT"
FT   Misc-difference     39
FT                       /note= "encoded by ATC"
FT   Misc-difference     40
FT                       /note= "encoded by TTT"
FT   Misc-difference     41
FT                       /note= "encoded by AGG"
FT   Misc-difference     42
FT                       /note= "encoded by TGT"
FT   Misc-difference     43
FT                       /note= "encoded by AAC"
FT   Misc-difference     424
FT                       /note= "encoded by TAT"
FT   Misc-difference     426
FT                       /note= "encoded by TCC"
FT   Misc-difference     428
FT                       /note= "encoded by GAC"
FT   Misc-difference     429
FT                       /note= "encoded by AGC"
FT   Misc-difference     434
FT                       /note= "encoded by AAA"
FT   Misc-difference     435
FT                       /note= "encoded by ACA"
FT   Misc-difference     442
FT                       /note= "encoded by CCT"
FT   Misc-difference     444
FT                       /note= "encoded by GAT"
FT   Misc-difference     446
FT                       /note= "encoded by ACA"
FT   Misc-difference     448
FT                       /note= "encoded by GAA"
FT   Misc-difference     454
FT                       /note= "encoded by CGC"
XX
PN   WO9736914-A1.
XX
PD   09-OCT-1997.
XX
PF   01-APR-1997;   97WO-US04707.
XX
PR   01-APR-1996;   96US-0617697.
XX
PA   (BARE/) BARENKAMP S J.
XX
PI   Barenkamp SJ;
XX
DR   WPI; 1997-503038/46.
DR   N-PSDB; AAT90995.
XX
PT   High molecular weight proteins of non-typeable Haemophilus
PT   influenzae - useful for vaccine production
XX
PS   Claim 7; Page 73-78; 183pp; English.
XX
CC   This protein comprises the high molecular weight surface protein
CC   HMW2 (123 kDa) of non-typeable Haemophilus influenzae strain 12 that
CC   has the immunological ability to protect against disease caused by a
CC   non-typeable Haemophilus strain and is characterised by at least
CC   one surface-exposed B-cell epitope that is recognised by monoclonal
CC   antibody AD6.  The HMW2 amino acid sequence was deduced from the
CC   hmw2 gene sequence (see AAT90995 and AAT90997).  The expressed protein
CC   is truncated, starting at residue 442 of the full-length gene
CC   product.  HMW1 (see AAW30293), HMW3 (see AAW30291) and HMW4 (see
CC   AAW30292) have also been identified.  A conjugate comprising HMW2
CC   linked to an antigen, hapten or polysaccharide, and a synthetic
CC   peptide of 6-150 amino acids corresponding to at least protective
CC   epitope of HMW2 are also claimed.  HMW proteins, conjugates and
CC   peptides can be used in vaccines, as immunogens for preparation of
CC   antibodies and as antigens for detection of these antibodies.
XX
SQ   Sequence   1477 AA;


alignment_scores:
            Quality:  280.50              Length:   1388
              Ratio:  0.426                 Gaps:     61
 Percent Similarity:  47.478   Percent Identity: 19.380

alignment_block:
 US-09-303-518D-649 x AAW30294   ..

 Align seg 1/1  to: AAW30294  from: 1  to: 1477

     175 GAAAATAAAGGCAAGTTTGCAGTCGGGGCGAAAGATATTGAGGTTTACAA 224
          :::|||:::|||   ::::::|||:::::  :::|||::::::
     219 LysAsnGluGlyValIleSerValAsnGlyGlySerIleSerLeu..... 233

     225 CAAAAAAGGGGAGTTGGTCGGCAAATCAATGACAAAAGCC.......... 264
                       |||:::|||:::::::::|||   :::
     234 .............LeuAlaGlyGlnLysIleThrIleSerAspIleIleA 246

     265 ..CCGATGATTGATTTTTCTGTGGTGTCGCGTAACGGCGTGGCGGCATTG 312
            |||   |||   :::|||:::::::::   ::::::   |||:::
     246 snProThrIleThrTyrSerIleAlaAlaProGluAsnGluAlaValAsn 262

     313 GTGGGCGATCAATATATTGTGAGCGTGGCACATAACGGCGGCTATAACAA 362
          :::||||||         :::      :::||||||      ||
     263 LeuGlyAsp..............IlePheAlaLysGlyGlyAsnIleAs 274

     363 CGTT..........................GATTTGGTGCGGAAGGAA 385
         ||||                         ::::::|||::::::
     274 nValArgAlaAlaThrIleArgAsnGlnGlyLysLeuSerAlaAspSerV 291

     386 GAAATCCCGATCAACATCGTTTTACTTATAAAATTGTGAAACGGAATAAT 435
            :::   |||:::                  :::|||:::::::::
     291 alSerLysAspLysSerGlyAsnIleValLeuSerAlaLysGluGlyGlu 307

     436 TATAAAGCAGGG..............ACTAAAGGCCATCCTTATGGCGG 470
            :::   |||               :::   :::     |||||
     308 AlaGluIleGlyGlyValIleSerAlaGlnAsnGlnGlnAlaLysGlyGl 324

     471 CGATTATCATATGCCGCGTTTGCATAAATTTGTCACAGATGCAGAACCTG 520
         |       :::                   :::       :::      :
     324 yLysLeuMetIleThrGlyAspLysValThrLeuLysThrGlyAlaValI 341

     521 TTGAAATGACCAGTTATATGGATGGGCGGAAATATATCGATCAAAATAAT 570
         :::::::::::::          |||:::   |||:::      ::::::
     341 leAspLeuSerGlyLysGluGlyGlyGluThrTyrLeuGlyGlyAspGlu 357

     571 TACCCTGACCGTGTTCGTATTGGGGCAGGCAGGCAATATTGGCGATCTGA 620
                                |||   |||::::::      :::
     358 Arg.................GlyGluGlyLysAsnGlyIleGlnLeuAl 368

     621 TGAAGATGAGCCCAATAACCGCGAAAGTTCATATCATATTGCAAGT.... 666
           :::         ::::::   |||:::   ::::::::::::
     368 aLysLysThrSerLeuGluLysGlySerThrIleAsnValSerGlyLysG 385

     667 .............GCGTATTCTTGG...........CTCGTTGGTGGC 690
                       |||      |||           |||:::   |||
     385 luLysGlyGlyArgAlaIleValTrpGlyAspIleAlaLeuIleAspGly 401

     691 AATACCTTTGCACAAAATGGATCAGGTGGTGGCACAGTCAACTTAGGTAG 740
         |||      ||||||   |||||||||
     402 AsnIleAsnAlaGln...GlySerGly...................... 409

     741 TGAAAAAATTAAACATAGCCCATATGGTTTTTTACCAACAGGAGGCTCAT 790
```

```
                                    :::      ||||||||
410 ........................AspIleAlaLysThrGlyGlyPheV 418

791 TTGGCGACAGTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCAAAAG 840
             ||||||     :::||||||   |||
418 alGluThrSerGlyHisAspLeuPheIleLysAsp.............. 429

841 TGGTTAATTAATGGGGTATTGCAAACGGGCAACCCCTATATAGGAAAAAG 890
        |||::::::::::::::
430 .........AsnAlaIleValAspAla...................... 435

891 CAATGGCTTCCAGCTGGTTCGTAAAGATTGGTTCTATGATGAAATCTTTG 940
                          |||:::|||:::   |||
436 .....................LysGluTrpLeuLeuAsp.......... 441

941 CTGGAGATACCCATTCAGTATTCTACGAACCACGTCAAAATGGGAAATAC 990
       |||:::  |||:::     |||          ::::::
442 ..PheAspAsnValSerIleAsnAlaGluAspProLeuPheAsnAsnThr 457

991 TCTTTTAACGACGAT...AATAATGGCACAGGAAAAATCAATGCCAAACA 1037
    ::::::|||||||:::     :::|||||||||:::   :::
458 GlyIleAsnAspGluPheProThrGlyThrGlyGluAlaSerAspProLy 474

1038 TGAACACAATTCTCTGCCTAATAGATTAAAAACACGAACCGTTCAATTGT 1087
     :::::::::|||    :::   |||   :::   |||::::::   :
474 sLysAsnSerGluLeuLysThrThrLeuThrAsnThrThrIleSerAsnT 491

1088 TT.....................AATGTTTCTTTATCCGAGACAGCAAGA 1116
     ::                     |||::::::   |||:::
491 yrLeuLysAsnAlaTrpThrMetAsnIleThrAlaSerArgLysLeuThr 507

1117 GAACCTGTTTATCATGCTGCAGGTGGTGTCAACAGTTATCGACCCAGACT 1166
                               |||||||||         ::
508 ..........................ValAsnSer.........SerIl 512

1167 GAATAATGGAGAAAATATTTCCTTTATTGACGAAGGAAAAGGCGAATTGA 1216
     :|||   |||:::|||       :::|||    ::::::||||||
512 eAsnIleGlySerAsnSerHisLeuIleLeuHisSerLysGly....... 526

1217 TACTTACCAGCAACATCAATCAAGGTGCTGGAGGATTATATTTCCAAGGA 1266
                              |||   :::|||||||:::   :::::|||
527 ....................GlnArgGlyGlyGlyValGlnIleAspGly 536

1267 GATTTACGGTCTCGCCTGAAAATAACGAAACTTGGCAAGGCGCGGGCGT 1316
     |||:::|||   :::      |||         ::::::|||:::   ||
537 AspIleThrSerLysGlyGlyAsnLeuThrIleTyrSerGlyGlyTrpVa 553

1317 TCATATCAGTGAAGACAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAA 1366
     |   :::   ::::::   |||:::
553 lAspValHisLysAsnIleThrLeu........................ 561

1367 ACGACCGCCTGTCCAAAATCGGCAAAGGCACGCTGCACGTTCAAGCCAAA 1416

561 .................................................. 561

1417 GGGGAAAACCAAGGCTCGATCAGCGTGGGCGACGGTACAGTCATTTTGGA 1466
          :::||||||   :::::::::      ::::::|||   :::
562 ......AspGlnGlyPheLeuAsnIleThrAlaAlaSerValAlaPhe.. 575

1467 TCAGCAGGCAGACGATAAAGGCAAAAAACAAGCCTTTAGTGAAATCGGCT 1516
      :::   ::::::::|||::::::       |||   ::::::|||
576 .GluGlyGlyAsnAsnLysAlaArgAspAlaAlaAsnAlaLysIle.... 590

1517 TGGTCAGCGGCAGGGGTACGGTGCAACTGAATGCCGATAATCAGTTCAAC 1566
          ::::::|||||||||   ::::::::::::::::::
591 .....ValAlaGlnGlyThrValThrIleThrGlyGluGlyLys...... 603

1567 CCCGACAAACTCTATTTCGGCTTTCGCGGCGGACGTTTGGATTTAAACGG 1616
                       ||||||:::::::::::::::||||||||
604 ..................AspPheArgAlaAsnAsnValSerLeuAsnGl 614

1617 GCATTCGCTTTCGTTCCACCGTATTCAAAATACCGATGAAGGGGCGATGA 1666
     |                              |||   :::|||      |

614 y...........................ThrGlyLysGlyLeuAsnI 621

1667 TTGTC...........AACCACAATCAAGACAAAGAATCCACCGTTACC 1704
     ||:::           |||   :::::::::   ::::::|||::::::
621 leIleSerSerValAsnAsnLeuThrHisAsnLeuSerGlyThrIleAsn 637

1705 ATTACAGGCAATAAAGATATTGCTACAACCGGCAATAACAACAGCTTGGA 1754
     |||:::||||||      |||      |||
638 IleSerGlyAsnIleThrIleAsnGlnThr................... 647

1755 TAGCAAAAAAGAAATTGCCTACAACGGTTGGTTTGGCGAGAAAGATACG. 1803
      ::::::|||:::   :::|||       |||        :::   |||:::
648 .ThrArgLysAsnThrSerTyr......TrpGlnThrSerHisAspSerH 662

1804 ......................ACCAAAACGAACGGGCGGCTCAACCTT 1830
                           :::|||:::::::::::::::::::
662 isTrpAsnValSerAlaLeuAsnLeuGluThrGlyAlaAsnPheThrPhe 678

1831 GTTTACCAGCCCGCCGCAGAAGACCGCACCCTGCTGCTT........... 1869
     :::          ::::::::::::::   |||
679 IleLysTyrIleSerSerAsnSerLysGlyLeuThrThrGlnTyrArgSe 695

1870 .TCCGGCGGAACAAATTTAAACGGCAACATCACGCAAACAAACGGCAAAC 1918
      |||:::|||:::|||:::||||||              :::||||||:::
695 rSerAlaGlyValAsnPheAsnGly............ValAsnGlyAsnM 708

1919 TGTTTTTCAGCGGCAGACCAACACCGCACGCCTACAATCATTTAAACGAC 1968
     ::   |||
708 etSerPhe......................................... 710

1969 CATTGGTCGCAAAAAGAGGGCATTCCTCGCGGGGAAATCGTGTGGGACAA 2018
          :::   |||||||||         :::      :::      ||
711 ......AsnLeuLysGluGlyAlaLysValAsnPheLysLeuLysProAs 725

2019 CGACTGGATCAACCGCACA.................TTTAAAGCGGAAA 2050
     |:::   :::|||   :::                      |||   |||   |
725 nGluAsnMetAsnThrSerLysProLeuProIleArgPheLeuAla...A 741

2051 ACTTCCAAATTAAAGGCGGACAGGCGGTGGTT................. 2082
     ||:::            ||||||   :::|||
741 snIleThrAlaThrGlyGlyGlySerValPhePheAspIleTyrAlaAsn 757

2083 ...TCCCGCAATGTTGCCAAAGTGAAA...GGCGATTGGCATTTGAGCAA 2126
        |||   :::   |||::::::|||   ::::::   ::::::||||||
758 HisSerGlyArgGlyAlaGluLeuLysMetSerGluIleAsnIleSerAs 774

2127 TCACGCCCAAGCAGTTTTTGGTGTCGCACCGCATCAAAGCCACACAATCT 2176
     |   |||:::   ::::::               :::||||||
774 nGlyAlaAsnPheThrLeu...............AsnSerHis....... 783

2177 GTACACGTTCGGACTGGACGGGTCTGACAAATTGTGTCGAAAAAACCATT 2226
       :::|||:::|||   :::         |||      :::   ||||||
784 ..ValArgGlyAspAspAlaPheLysIleAsn...LysAspLeuThrIle 798

2227 ACCGACGATAAAGTGATTGCTTCATTGACTAAGACCGACATCAGCGGCAA 2276
     :::      :::         ||||||   :::|||      :::
799 AsnAlaThrAsnSerAsnPheSerLeuArgGlnThrLysAspAspPheTy 815

2277 TGTCGATCTTGCCGATCACGCTCATTTAAATCTCACAGGGCTTGCCACAC 2326
           |||   :::|||      :::      :::::::::   |
815 rAspGlyTyrAlaArgAsnAlaIleAsnSerThrTyrAsnIleSerIleL 832

2327 TCAACGGCAATCTTAGTGCAAATGGCGATACACGTTATACAGTCAGCCAC 2376
     ||:::|||||||::::::   :::|||
832 euGlyGlyAsnValThrLeuGlyGly....................... 840

2377 AACGCCACCCAAAACGGCAACCTTAGCCTCGTGGGCAATGCCCAAGCAAC 2426
              ||||||::::::   |||:::::::||||||      ||
841 .........GlnAsnSerSerSerSerIleThrGlyAsn......IleTh 852

2427 ATTTAATCAAGCC.........ACATTAAACGGCAACACATCGGCTTCGG 2467
     |:::::::::|||         ||||||:::::::|||::::::   :::
852 rIleGluLysAlaAlaAsnValThrLeuGluAlaAsnAsnAlaProAsnG 869
```

```
2468 GCAATGCTTCATTTAATCTAAGCGACCACGCCGTACAAAACGGCAGTCTG 2517
       :::         |||:::   |||::::::::::::    |||||||||
 869 lnGln.........AsnIleArgAspArgValIleLysLeuGlySerLeu 882

2518 ACGCTTTCCGGCAACGCTAAGGCAAACGTAAGCCATTCCGCACTCAACGG 2567
        ::::::|||:::   :::   :::   :::::::::   :::::::||
 883 LeuValAsnGlySerLeuSerLeuThrGlyGluAsnAlaAspIleLysGl 899

2568 TAATGTCTCCCTAGCCGATAAGGCAGTATTCCAT.........TTTGAAA 2608
     ||||::::::::::::::::::|||:::|||               ::::
 899 yAsnLeuThrIleSerGluSerAlaThrPheLysGlyLysThrArgAspT 916

2609 GCAGCCGCTTTACCGGACAAATCAGCGGCGGCAAGGATACGGCATTACAC 2658
     ::   ::::::|||||::::::::::::::::      :::   ::::::
 916 hrLeuAsnIleThrGlyAsnPheThrAsnAsnGlyThrAlaGluIleAsn 932

2659 TTAAAAGACAGCGAATGGACGCTGCCGTCAGGCACGGAATTAGGCAATTT 2708
     :::   ::::::            |||   :::   |||:::   |||:::||
 933 IleThrGlnGlyValValLysLeuGlyAsnValThrAsnAspGlyAspLe 949

2709 AAACCTTGACAACGCCACCATTACACTCAATTCCGCCTATCGCCACGATG 2758
     ||||:::   :::
 949 uAsnIleThrThr..................................... 953

2759 CGGCAGGGGCGCAAACCGGCAGTGCGACAGATGCGCCGCGCCGCCGTTCG 2808
                                               ::::::
 954 ..........................................HisAla 955

2809 CGCCGTTCGCGCCGTTCCCTATTATCCGT....TACACCGCCAACTTCGG 2854
     :::|||::::::|||||||::::::::               :::   ||
 956 LysArgAsnGlnArgSerIleIleGlyGlyAspIleIleAsnLysLysGl 972

2855 TAGAATCCCGTTTCAACACGCTGACGGTAAACGGCAAATTGAACGGTCAG 2904
     |    :::   :::      :::::::::      :::|||:::    :::|
 972 ySerLeuAsnIleThrAspSerAsnAsnAspAlaGluIleGlnIleGlyG 989

2905 GGAACATTCCGCTTTATGTCGGAACTCTTCGGCTACCGCAGCGACAAATT 2954
     ||||||||             |||:::|||    :::      |||||||||::
 989 lyAsnIleSerGlnLysGluGlyAsnLeuThrIle.SerSerAspLysIl 1005

2955 GAAGCTGGCGGAAAGTTCCGAAGGCACTTACACCTTGGCGGTCAACAATA 3004
     :::::::::
1005 eAsnIleThr........................................ 1008

3005 CCGGCAACGAACCTGCAAGCCTCGAACAATTGACGGTAGTGGAAGGAAAA 3054
                             :::|||:::|||:::   :::|||
1009 ......................LysGlnIleThrIleLysLysGlyIle 1017

3055 GACAACAAACCGCTGTCC...........GAAAACCTTAATTTCACCCT 3092
     |||:::::        |||             :::|||    |||:::|||::
1018 AspGlyGluAspSerSerSerAspAlaThrSerAsnAlaAsnLeuThrIl 1034

3093 GCAAAACGAACACGTCGAT....................GCCGGCGCGT 3121
     :::::::::::::::                           :::|||
1034 eLysThrLysGluLeuLysLeuThrGluAspLeuSerIleSerGlyPheA 1051

3122 GGCGTTACCAACTCATCCGCAAAGACGGC...GAGTTCCGCCTGCATAAT 3168
        :::    ::::::      |||||||||    ::::::   :::    |||
1051 snLysAlaGluIleThrAlaLysAspGlyArgAspLeuThrIleGlyAsn 1067

3169 CCGGTCAAAGAACAAGAGCTTTCCGACAAACTCGGCAAGGCAGAAGCCAA 3218
                              :::|||       :::     |||||||||
1068 Ser.................AsnAspGlyAsnSerGlyAlaGluAla.. 1077

3219 AAAACAGGCGGAAAAAGACAACGCGCAAAGCCTTGACGCGCTGATTGCGG 3268
       |||    :::        :::|||::::::      |||       |||:::|
1078 .LysThrValThrPheAsnAsnValLys......AspGlnLysIleSerA 1092

3269 CCGGGCGCGATGCCGTCGAAAAGACAGAAAGCGTTGCCGAACCGGCCCGG 3318
     ||               |||      :::::::::|||         :::
1092 laAspGlyHisAsnValThrLeuAsnSerLysValLysThrSerSerSer 1108

3319 CAGGCAGGCGGGGAAAATGTCGGCATTATGCAGGCGGAGGAAGAGAAAAA 3368
     ::::::|||   |||:::
1109 AsnGlyGlyArgGluSer............................... 1114

3369 ACGGGTGCAGGCGGATAAAGACACCGCCTTGGCGAAACAGCGCGAAGCGG 3418
                   ::::::|||:::||||||:::|||
1115 .......AsnSerAspAsnAspThrGlyLeu.................. 1122

3419 AAACCCGGCCGGCTACCACCGCCTTCCCCCGCGCCCGCCGCGCCCGCCGG 3468
                                 |||          :::        ::::::::::::
1123 ................ThrIleThrAlaLysAsnValGluValAsnLys 1133

3469 GATTTGCCGCAACTGCAACCCCAACCGCAGCCCCAACCGCAGCGCGACCT 3518
     |||:::   :::|||
1134 AspIleThrSerLeu................................. 1138

3519 GATCAGCCGTTATGCCAATAGCGGTTTGAGTGAATTTTCCGCCACGCTCA 3568
             :::   :::|||           ||||||      :::|||  :::
1139 .......LysThrValAsnIleThrAlaSerGluLysValThrThr.Thr 1152

3569 ACAGCGTTTTCGCCGTACAGGACGAATTAGACCGCGTATTTGCCGAAGAC 3618
     ::::::   :::            ||||||      |||   :::
1153 AlaGlySerThrIleAsnAlaThrAsnGlyLysAlaSerIleThrThrLy 1169

3619 CGCCGCAACGCCGTTTGGACAAGCGGCATCCGGGACACCAAACACTACCG 3668
      ::::::            |||        :::::::::|||    :::::::::
1169 sThrGlyAspIleSerGlyThrIleSerGlyAsnThrValSerValSerA 1186

3669 TTCGC...AAGATTTCCGCGCCTACCGCCAACAAACCGACCTGCGCCAAA 3715
     ::             :::::::::   :::::::::|||        :::   :::
1186 laThrGlyAspLeuThrThrLysSerGlySerLysIleGluAlaLysSer 1202

3716 TCGGTATGCAGAAAAACCTCGGCAGCGGGCGCGTCGGCATCCTGTTTTCG 3765
     :::   ::::::    |||||||||:::|||:::    ::::::
1203 GlyGluAlaAsnValThrSerAlaThrGlyThrIleGlyGly........ 1216

3766 CACAACCGGACCGAAAACACCTTCGACGACGGCATCGGCA.......... 3805
       |||   :::    :::|||    ::::::|||||||:::|||
1217 .ThrIleSerGlyAsnThrValAsnValThrAlaAsnAlaGlyAspLeuT 1233

3806 ........................ACTCGGCACGGCTTGCCCACGGCG 3829
                                :::   :::|||      |||
1233 hrValGlyAsnGlyAlaGluIleAsnAlaThrGluGlyAlaAlaThrLeu 1249

3830 CCGTTTTCGGGCAATACGGCATCGACA......GGTTCTACATCGGCATC 3873
           :::|||||||||    :::|||      |||||||:::    :::||
1250 ThrAlaThrGlyAsnThrLeuThrThrGluAlaGlySerSerIleThrSe 1266

3874 AGCGCGGGCGCGGGTTTTAGCAGCGGCAGCCTTTCAGACGGCATCGGAGG 3923
     |::::::::::::|||            |||      |||::::::|||   |
1266 rThrLysGlyGlnValAspLeuLeuAla......GlnAsnGlySerIleA 1281

3924 CAAAATCCGCCGCCGCGTGCTGCATTACGGCATTCAGGCACGATACCGCG 3973
     ||   |||         |||:::   :::|||                  |||
1281 laGlySerIleAsnAlaAlaAsnValThrLeuAsnThrThrGlyThrLeu 1297

3974 CCGGTTTCGGCGGATTCGGCATCGAACCGCACATCGGCGCAACGCGCTAT 4023
        :::   |||   |||      :::    |||||||:::        :::||
1298 ThrThrValAlaGlySerAspIleLysAlaThrSerGlyThrLeuValIl 1314

4024 TTCGTCCAAA 4033
     |:::::::|||
1314 eAsnAlaLys 1317
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:AAR63505

seq_documentation_block:

```
ID   AAR63505 standard; Protein; 1536 AA.
XX
AC   AAR63505;
XX
DT   25-JUN-1995  (first entry)
XX
```

```
DE   Haemophilus high molecular weight protein HMW1.
XX
KW   High molecular weight protein; HMW1; protective vaccine; otitis;
KW   sinusitis; bronchitis; Hib.
XX
OS   Haemophilus.
XX
PN   WO9421290-A.
XX
PD   29-SEP-1994.
XX
PF   15-MAR-1994;   94WO-US02550.
XX
PR   16-MAR-1993;   93US-0038682.
XX
PA   (BARE/) BARENKAMP S J.
PA   (SGEM/) ST GEME J W.
XX
PI   Barenkamp SJ,  St GEME JW;
XX
DR   WPI; 1994-316665/39.
DR   Q-PSDB; Q72293.
XX
PT   New immunogenic high mol. wt. proteins of non typeable
PT   Haemophilus - useful in protective vaccines
XX
PS    Claim 2; Page 31; 127pp; English.
XX
CC   The HMW1 protein encoded by this sequence is useful in a vaccine to
CC   protect against disease caused by non-typeable Haemophilus which are
CC   not controlled by H. influenzae type b (Hib) vaccines.  The encoded
CC   protein can also be used as a carrier for protective Hib
CC   polysaccharide (in a conjugate vaccin against meningitis) or for
CC   other antigens, haptens, etc.
XX
SQ   Sequence   1536 AA;

alignment_scores:
          Quality:  277.50            Length:    1182
            Ratio:  0.460               Gaps:      60
 Percent Similarity: 51.015   Percent Identity: 20.981

alignment_block:
 US-09-303-518D-649 x AAR63505   ..

 Align seg 1/1  to: AAR63505  from: 1  to: 1536

    127 GGACACACTTATTTCGGCATCAACTACCAATACTATCGCGACTTTGCCGA 176
        |||:::||||||:::|||                           :::::
    349 GlyGluThrTyrLeuGly..........................GlyAs 356

    177 AAATAAAGGCAAGTTTGCAGTCGGGGCGAAAGATATTGAGGTTTACAACA 226
        :::::::|||:::           |||   |||   |||::::::   :::|
    356 pGluArgGlyGlu.........GlyLysLysGlyIleGlnLeuAlaLysL 370

    227 AAAAAGGGGAGTTGGTCGGCAAATCAATGACAAAAGCCCCGATGATTGAT 276
        ||                  |||:::   |||:::      |||:::
    370 ys..................ThrSerLeuGluLysGlySerThrIleAsn 380

    277 TTTTCTGTGGTGTCGCGTAACGGCGTGGCGGCATTGGTGGGCGATCAATA 326
          |||      :::::::::|||   |||   :::   ||||||
    381 ValSerGlyLysGluLysGlyGlyArgAlaIleValTrpGlyAspIleAl 397

    327 TATTGTG..............................AGCGTGGCACATA 346
          ::::::                              :::::::|||   :
    397 aLeuIleAspGlyAsnIleAsnAlaGlnGlySerGlyAspIleAlaLysT 414

    347 ACGGCGGCTAT....................................... 357
         ::|||||||:::
    414 hrGlyGlyPheValGluThrSerGlyHisAspLeuPheIleLysAspAsn 430

    358 ...................................AACAACGTTGATTT 371
                                        :::|||||||::::
    431 AlaIleValAspAlaLysGluTrpLeuLeuAspProAspAsnValSerIl 447

    372 TGGTGCGGAA......GGAAGAAATCCCGATCAACATCGTTTACTTAT. 414
        :::||||||       ||||||:::      ::::::       |||
    447 eAsnAlaGluThrAlaGlyArgSerAsnThrSerGluAspAspGluTyrT 464

    415 ....................AAAATTGTGAAACGGAATAATTATAAAGCA 444
                            :::     |||||||||:::    |||:::
    464 hrGlySerGlyAsnSerAlaSerThrProLysArgAsnLysGluLysThr 480

    445 GGGACTAAAGGCCATCCTTATGGCGGCGATTATCATATGCCGCGTTTGCA 494
           |||          :::   :::|||:::          :::    :
    481 .ThrLeuThrAsnThrThrLeuGluSerIleLeuLysLysGlyThrPheV 497

    495 TAAATTTGTCACAGATGCAGAACCTGTTGAAATGACCAGTTATATGGATG 544
        ::|||::::::      |||:::                   |||::::::
    497 alAsnIleThrAlaAsnGlnArg..................IleTyrVal 507

    545 GGCGGAAATATATCGATCAAAATAATTACCCTGACCGTGTTCGTATTGGG 594
        :::::::::|||:::::::::         :::||||||:::::      ||
    508 AsnSerSerIleAsnLeuSerAsnGlySerLeuThrLeuTrpSerGluGl 524

    595 GCAGGCAGGC...........................AATATTGGCGATC 617
        |::::::|||                           :::    ||||||
    524 yArgSerGlyGlyGlyValGluIleAsnAsnAspIleThrThrGlyAspA 541

    618 TGATGAAGATGAGCCCAATAACC..................GCGAAAGTT 649
             :::    :::    :::|||                  :::    |||
    541 spThrArgGlyAlaAsnLeuThrIleTyrSerGlyGlyTrpValAspVal 557

    650 CATATCATATTGCAAGTGCGTATTCTTGGCTCGTTGGTGGCAATACCTTT 699
        |||      :::::::::          |||:::::::   :::      |||
    558 HisLysAsnIleSerLeuGlyAlaGlnGlyAsnIle..AsnIleThrAla 573

    700 GCACAAAAT...............GGATCA...............GGTGG 719
           |||:::               ||||||                |||
    574 LysGlnAspIleAlaPheGluLysGlySerAsnGlnValIleThrGlyGl 590

    720 TGGCACAGTCAACTTAGGTAGTGAAAAAATTAAACATAGCCCATATGGTT 769
          ||||||:::::    |||::::::|||                  ||||
    590 nGlyThrIleThrSerGlyAsnGlnLys..................GlyP 601

    770 TTTTACCAACAGGAGGCTCATTTGGCGACAGTGGCTCACCAATGTTTATC 819
        ||          ::::::    |||::::::    :::||||||    :::     :::
    601 heArgPheAsnAsnValSerLeuAsnGlyThrGlySerGlyLeuGlnPhe 617

    820 TATGATGCCCAAAAGCAAAAGTGGTTAATTAATGGGGTATTGCAAACGGG 869
                :::    :::|||:::    |||::::::      ::::::
    618 ThrThrLysArgThrAsnLysTyrAlaIleThrAsnLysPheGluGlyTh 634

    870 CAACCCCTATATAGGAAAAAGCAAT...GGCTTCCAGCTGGTTCGTAAAG 916
                      ||||||    |||    ::::::    |||    :::::::
    634 rLeuAsnIleSerGlyLysValAsnIleSerMetValLeuProLysAsnG 651

    917 ATTGGTTCTATGATGAAATCTTTGCTGGAGATACCCAT......TCAGTA 960
        ::          |||||||:::    |||    |||    |||:::        :::
    651 luSerGlyTyrAspLys...PheLysGlyArgThrTyrTrpAsnLeuThr 666

    961 TTCTACGAACCACGTCAAAATGGGAAATACTCTTTTAACGACGATAATAA 1010
               :::      ::::::|||:::::::::::::::    ||||:::::
    667 SerLeuAsnValSerGluSerGlyGluPheAsnLeuThrIleAspSerAr 683

   1011 TGGCACA.........GGAAAAATCAATGCCAAACATGAACACAATTCTC 1051
        :|||:::            |||    ::::::      ::::::    |||::::
    683 gGlySerAspSerAlaGlyThrLeuThrGlnProTyrAsnLeuAsnGlyI 700

   1052 TGCCT...AATAGATTAAAAACACGAACCGTTCAATTGTTTAATGTTTCT 1098
        ::        |||:::         |||    :::|||:::       |||:::
    700 leSerPheAsnLysAspThrThrPheAsnValGluArg...AsnAlaArg 715

   1099 TTATCCGAGACAGCAAGAGAACCTGTTTATCATGCTGCAGGTGGTGTCAA 1148
        ::::::            :::    |||:::                   |||:::||
```

```
 716 ValAsnPheAspIleLysAlaProIle..............GlyIleAs 727

1149 CAGTTATCGACCCAGACTGAATAATGGAGAAAATATTTCCTTTATTGACG 1198
     |:::|||                          ::::::::::::::   :::
 727 nLysTyr....................SerSerLeuAsnTyrAlaSer. 736

1199 AAGGAAAAGGCGAATTGATACTTACCAGCAACATCAATCAAGGTGCTGGA 1248
                                ::::::::||||||:::   ::::::|||
 737 ...................PheAsnGlyAsnIleSerValSerGlyGly 746

1249 GGATTATATTTCCAAGGAGATTTTACGGTCTCGCCTGAAAATAACGAAAC 1298
     |||          :::  ||||||||:::       ::::::::::::::
 747 Gly.........SerValAspPheThrLeuLeuAlaSerSerSerAsnVa 760

1299 TTGGCAAGGCGCGGGCGTT...................CATATCAGTG 1327
     :   |||       |||||                       ::::::|||
 760 l...GlnThrProGlyValValIleAsnSerLysTyrPheAsnValSerT 776

1328 AAGACAGTACCGTTACTTGGAAAGTAAACGGC................. 1359
            |||:::::  :::|||::::::|||
 776 hrGlySerSerLeuArgPheLysThrSerGlySerThrLysThrGlyPhe 792

1360 ...GTGGCAAACGAC...CGCCTGTCCAAAATCGGCAAAGGCACGCTGCA 1403
        :::   :::|||         |||:::      |||   :::
 793 SerIleGluLysAspLeuThrLeuAsnAlaThrGlyGlyAsnIleThrLe 809

1404 CGTTCAAGCCAAAGGGGAAAAC.........................C 1426
       :::|||::::::|||   :::                          :
 809 uLeuGlnValGluGlyThrAspGlyMetIleGlyLysGlyIleValAlaL 826

1427 AAGGCTCGATCAGCGTGGGCGACGGTACAGTCATTTGGATCAGCAGGCA 1476
     ::    :::|||:::            |||::::::   :::
 826 ysLysAsnIleThrPheGluGlyGlyAsnIleThrPhe............ 838

1477 GACGATAAAGGCAAAAAACAAGCCTTTAGTGAAATCGGCTTGGTCAGCGG 1526
                |||::::::::|||   :::||||||          :::||
 839 .........GlySerArgLysAlaValThrGluIle.........GluGl 849

1527 CAGGGGTACGGTGCAACTGAATGCCGATAATCAGTTCAACCCCGACAAAC 1576
     |:::    |||::::::   ||||||:::
 849 yAsnValThrIleAsnAsnAsnAlaAsn..................... 858

1577 TCTATTTCGGCTTTCGCGGCGGACGTTTGGATTTAAACGGGCATTCGCTT 1626
                                :::    |||    |||    ::::::
 859 ........................ValThrLeuIleGlySerAspPhe 866

1627 TCGTTCCACCGTATTCAAAATACCGATGAAGGGGCGATGATTGTCAACCA 1676
     :::    |||:::            |||    :::        :::|||:::|||
 867 AspAsnHisGlnLysProLeuThrIleLysLysAspValIleIleAsnSe 883

1677 CAATCAA......GACAAAGAATCCACCGTTACCATTACAGGCAATAAAG 1720
       ::::::                 :::    |||:::|||:::||||||
 883 rGlyAsnLeuThrAlaGlyGlyAsnIleValAsnIleAlaGlyAsnLeuT 900

1721 ATATTGCTACAACCGGCAAT............................. 1740
       :::    :::::::|||
 900 hrValGluSerAsnAlaAsnPheLysAlaIleThrAsnPheThrPheAsn 916

1741 ...............AACAACAGCTTGGATAGCAAAAAAGAAATTGCCTA 1775
                     :::|||:::    :::|||:::    :::||||||
 917 ValGlyGlyLeuPheAspAsnLysGlyAsnSerAsnIleSerIleAlaLy 933

1776 CAACGGT...TGGTTTGGCGAGAAAGATACGACCAAAACGAACGGGCGGC 1822
       :::|||       |||    :::    |||::::::|||         :::|
 933 sGlyGlyAlaArgPheLysAspIleAspAsnSerLys.........AsnL 947

1823 TCAACCTTGTTTACCAGCCCGCCGCAGAAGACCGCACCCTGCTGCTTTCC 1872
     ||::::::::   :::    ::::::            ||||||:::::    :::
 947 euSerIleThrThrAsnSerSerSerThrTyrArgThrIleIleSerGly 963

1873 GGCGGAACAAATTTAAACGGCAACATCACGCAAACAAACGGCAAACTGTT 1922
     :::    ||||||    ||||||:::::::::    ||||||
 964 AsnIleThrAsnLysAsnGlyAspLeuAsnIleThrAsn........... 976

1923 TTTCAGCGGCAGACCAACACCGCACGCCTACAATCATTTAAACGACCATT 1972
          :::|||    |||                  :::::::::
 977 ....GluGlySerAspThrGlu.........MetGlnIleGlyGlyAspV 989

1973 GGTCGCAAAAAGAGGGCATTCCTCGCGGGGAAATCGTGTGGGACAACGAC 2022
       ||||||||||||                  |||:::::::::   ::::::|||
 989 alSerGlnLysGlu............GlyAsnLeuThrIleSerSerAsp 1001

2023 TGGATCAACCGCACATTTAAAGCGGAAAACTTCCAAATTAAAGGCGGACA 2072
       ||||||    |||              :::::::::    ||||||:::|||
1002 LysIleAsnIleThr.........LysGlnIleThrIleLysAlaGly.. 1014

2073 GGCGGTGGTTTCCCGCAATGTTGCCAAAGTGAAAGGCGATTGGCATTTGA 2122
                                |||         ::::::|||       :
1015 ..................ValAspGlyGluAsnSerAspSerAspAlaT 1025

2123 GCAATCACGCCCAAGCAGTTTTTGGTGTCGCACCGCATCAAAGCCACACA 2172
     ::|||:::|||:::                                      |||
1025 hrAsnAsnAlaAsnLeu............................Thr 1031

2173 ATCTGTACACGTTCGGACTGGACGGGTCTGACAAATTGTGTCGAAAAAAC 2222
     |||    |||:::                                   |||
1032 IleLysThrLys...........................GluLeuLy 1038

2223 CATTACCGACGATAAAGTGATTGCTTCATTGACTAAGACCGACATCAGCG 2272
       :::|||:::|||         |||::::::::::::|||::::::|||:::
1038 sLeuThrGlnAspLeuAsnIleSerGlyPheAsnLysAlaGluIleThrA 1055

2273 GCAATGTCGATCTTGCCGATCACGCTCATTTAAATCTCACAGGGCTTGCC 2322
     :::::    |||    :::|||
1055 laLys...AspGlySerAsp.........................Leu 1061

2323 ACACTCAACGGCAATCTTAGTGCAAATGGCGATACACGTTATACAGTCAG 2372
     |||::::::::::::    ||||||:::|||    :::          |||::
1062 ThrIleGlyAsnThrAsnSerAlaAspGlyThrAsnAlaLysLysValTh 1078

2373 CCACAACGCCACCCAAAACGGCAACCTTAGCCTCGTGGGCAATGCCCAAG 2422
     :    |||    ::::::::::::::::|||       |||
1078 rPheAsnGlnValLysAspSerLysIleSerAlaAspGly.......... 1091

2423 CAACATTTAATCAAGCCACATTAAACGGCAACACATCGGCTTCGGGCAAT 2472
                ::::::::||||||::::::::::::::::::||||||:::
1092 ........HisLysValThrLeuHisSerLysValGluThrSerGlySer 1105

2473 GCTTCATTTAATCTAAGCGACCACGCCGTACAAAACGGCAGTCTGACGCT 2522
         :::    |||    :::|||    :::    :::|||::::::||||||::
1106 AsnAsn...AsnThrGluAspSerSerAspAsnAsnAlaGlyLeuThrIl 1121

2523 TTCCGGCAACGCTAAGGCAAACGTAAGCCATTCCGCACTCAACGGTAATG 2572
     :       :::||||||   ||||||:::              :::|||:::|||:
1121 e......AspAlaLys...AsnValThr.........ValAsnAsnAsnI 1132

2573 TCTCCCTAGCCGATAAGGCAGTATTCCATTTTGAAAGCAGCCGCTTTACC 2622
     :::::    :::    |||||||||              :::    |||       :::
1132 leThr...SerHisLysAlaVal.........SerIleSerAlaThrSer 1144

2623 GGACAAATCAGCGGCGGCAAGGATACGGCATTACACTTAAAA......GA 2666
     |||:::|||:::                  |||:::::::::               ::
1145 GlyGluIleThrThrLysThrGlyThrThrIleAsnAlaThrThrGlyAs 1161

2667 CAGCGAATGGACGCTGCCGTCAGGCACGGAATTAGGCAATTTAAACCTTG 2716
     :    |||    |||         :::|||:::     ||||||::::::::    :
1161 nValGluIleThrAlaGlnThrGlySerIleLeuGlyGlyIleGluSerS 1178

2717 ACAACGCCACCATTACACTCAATTCCGCCTATCGCCACGATGCGGCAGGG 2766
     :::::::::::::||||||:::::::::                 |||::::::
1178 erSerGlySerValThrLeuThrAlaThrGluGlyAlaLeuAlaValSer 1194

2767 GCGCAAACCGGCAGTGCGACAGATGCGCCGCGCCGCCGTTCGCGCCGTTC 2816
           :::|||:::::::::   :::          :::    :::       ::
1195 AsnIleSerGlyAsnThrValThrValThrAlaAsnSerGlyAlaLeuTh 1211
```

```
2817 GCGCCGTTCCCTATTATCCGTTACACCGCCAACTTCGGTAGAATCCCGTT 2866
      :     :::     ::::::                 ||||||   :::
1211 rThrLeuAlaGlySerThrIleLysGlyThrGluSerValThrThrSerS 1228

2867 TCAACACGCTGACGGTAAACGGCAAATTGAACGGTCAGGGAACA...... 2910
       ::::::       :::::|||    ::::::|||   ||||||
1228 erGlnSerGlyAspIleGlyGlyThrIleSerGly...GlyThrValGlu 1243

2911 TTCCGCTTTATGTCGGAACTCTTCGGCTACCGCAGCGACAAATTGAAGCT 2960
       :::      :::::|||                :::::|||:::|||
1244 ValLysAlaThrGluSerLeuThrThrGlnSerAsnSerLysIleLys.. 1259

2961 GGCGGAAAGTTCCGAAGGCACTTACACCTTGGCGGTC............ 2997
      |||   ::::::|||::::::   |||    |||:::
1260 .AlaThrThrGlyGluAlaAsnValThrSerAlaThrGlyThrIleGlyG 1276

2998 ...........AACAATACCGGCAACGAACCTGCAAGCCTCGAACAATTG 3036
                :::||||||   |||      |||:::     :::|||
1276 lyThrIleSerGlyAsnThrValAsnValThrAlaAsnAlaGlyAspLeu 1292

3037 ACGGTA......................GTGGAAGGAAAAGACAACAA 3062
     ||||||                                   :::
1293 ThrValGlyAsnGlyAlaGluIleAsnAlaThrGluGlyAlaAlaThrLe 1309

3063 ACCGCTGTCCGAAAACCTTAATTTCACCCTGCAAAACGAACACGTCGATG 3112
            |||::::::          |||       ::::::|||:::   :
1309 uThrThrSerSerGlyLysLeuThrThrGluAlaSerSerHisIleThrS 1326

3113 CCGGCGCGTGGCGTTACCAACTCATCCGCAAAGACGGCGAGTTCCGCCTG 3162
      :: |||:::|||    :::|||    ||||||:::    :::|||   |||
1326 er.AlaLysGlyGlnValAsnLeuSerAlaGlnAspGlySerValAlaGl 1342

3163 CATAATCCGGTCAAAGAACAAGAGCTTTCCGACAAACTCGGCAAGGCAGA 3212
           |||::::::   |||    ::::::    |||:::::::::
1342 ySerIleAsnAlaAlaAsnValThrLeuAsnThrThrGlyThrLeuThrT 1359

3213 AGCCAAAAAACAGGCGGAAAAAGACAACGCGCAAAGCCTTG 3253
           |||:::    :::    ::::::|||       :::|||
1359 hrValLysGlySerAsnIleAsnAlaThrSerGlyThrLeu 1372
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01846

```
seq_documentation_block:
ID   AAB01846 standard; Protein; 1536 AA.
XX
AC   AAB01846;
XX
DT   11-SEP-2000  (first entry)
XX
DE   Haemophilus influenzae strain 12 HMW1A protein, SEQ ID NO:67.
XX
KW   HMW protein; hmw gene; hmwA1; hmwA2; high molecular weight;
KW   non-typeable Haemophilus influenzae; NTHi; non-encapsulated;
KW   recombinant production; Escherichia coli; antibacterial; vaccine;
KW   human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;
KW   detection; diagnosis.
XX
OS   Haemophilus influenzae strain 12.
XX
PN   WO200020609-A2.
XX
PD   13-APR-2000.
XX
PF   07-OCT-1999;   99WO-CA00938.
XX
PR   07-OCT-1998;   98US-0167568.
PR   08-DEC-1998;   98US-0206942.
XX
PA   (CONN-) CONNAUGHT LAB LTD.
XX
PI   Loosmore SM,  Yang Y,  Klein MH;
XX
DR   WPI; 2000-303789/26.
DR   N-PSDB; AAA52195.
XX
PT   Nucleic acid molecule for producing recombinant high molecular weight
PT   proteins of Haemophilus which are used as a vaccine to provide
PT   protection against Haemophilus induced diseases in humans -
XX
PS   Example 16; Fig 28A-Q; 307pp; English.
XX
CC   The invention relates to the recombinant production of Haemophilus
CC   influenzae high molecular weight (HMW) proteins in Escherichia coli. The
CC   expression construct used to effect recombinant expression comprises a
CC   promoter functional in E. coli (e.g., the T7 promoter) operably linked
CC   to a modified hmwABC operon from a non-typeable (non-encapsulated) H.
CC   influenzae (NTHi). Most HMW-expressing NTHi strains contain two hmw gene
CC   clusters termed hmw1ABC and hmw2ABC. Each hmwABC operon comprises hmwA,
CC   hmwB and hmwC genes. The hmwA genes encode the structural HMWA proteins
CC   and the hmwB and hmwC genes encode accessory proteins which are
CC   responsible for post-translational processing and secretion of the HMWA
CC   proteins. The modified hmwABC operon used in the expression construct of
CC   the invention contains an A gene modified such that it encodes only the
CC   mature HMWA. The invention also discloses hmwA genes (AAA52175-A52198)
CC   and HMWA proteins (AAB01824-B01849) from the non-typeable H. influenzae
CC   strains Joyc, K1, K21, LCDC2, PMH1, 15 and 12. The nucleic acids and
CC   vectors are used for the production of recombinant H. influenzae HMW
CC   proteins which can be used as vaccines to mediate a humoral or
CC   cell-mediated immune response to provide protection against diseases in
CC   humans caused by H. influenzae (e.g., otitis media, epiglottitis,
CC   pneumonia and tracheobronchitis). The HMW proteins are also useful as
CC   antigens in immunoassays for detecting antibodies against Haemophilus,
CC   HMW proteins and/or HMW peptides. The nucleotide sequences encoding the
CC   HMW proteins can be used to isolate and clone hmw genes from other
CC   non-typeable strains of Haemophilus via hybridisation reactions. The
CC   present sequence represents an HMWA protein from a non-typeable strain of
CC   H. influenzae.
XX
SQ   Sequence   1536 AA;

alignment_scores:
           Quality:  277.50            Length:   1182
             Ratio:  0.460               Gaps:     60
 Percent Similarity:  51.015   Percent Identity: 20.981

alignment_block:
 US-09-303-518D-649 x AAB01846   ..

 Align seg 1/1  to: AAB01846  from: 1  to: 1536

 127 GGACACACTTATTTCGGCATCAACTACCAATACTATCGCGACTTTGCCGA 176
      |||:::||||||:::|||                          :::::
 349 GlyGluThrTyrLeuGly..........................GlyAs 356

 177 AAATAAAGGCAAGTTTGCAGTCGGGGCGAAAGATATTGAGGTTTACAACA 226
      ::::::|||:::             |||   |||   |||::::::   :::|
 356 pGluArgGlyGlu.........GlyLysLysGlyIleGlnLeuAlaLysL 370

 227 AAAAAGGGGAGTTGGTCGGCAAATCAATGACAAAAGCCCCGATGATTGAT 276
      ||                   |||:::   |||:::       |||:::
 370 ys.................ThrSerLeuGluLysGlySerThrIleAsn 380

 277 TTTTCTGTGGTGTCGCGTAACGGCGTGGCGGCATTGGTGGGCGATCAATA 326
         |||       :::::::::|||   |||    :::    ||||||
 381 ValSerGlyLysGluLysGlyGlyArgAlaIleValTrpGlyAspIleAl 397

 327 TATTGTG.............................AGCGTGGCACATA 346
      ::::::                              ::::::|||    :
 397 aLeuIleAspGlyAsnIleAsnAlaGlnGlySerGlyAspIleAlaLysT 414

 347 ACGGCGGCTAT................................... 357
      ::|||||||:::
 414 hrGlyGlyPheValGluThrSerGlyHisAspLeuPheIleLysAspAsn 430

 358 .................................AACAACGTTGATTT 371
```

```
                                                  :::|||||||::::
 431 AlaIleValAspAlaLysGluTrpLeuLeuAspProAspAsnValSerIl 447

 372 TGGTGCGGAA......GGAAGAAATCCCGATCAACATCGTTTTACTTAT. 414
     ::::||||||      ||||||:::     ::::::      |||
 447 eAsnAlaGluThrAlaGlyArgSerAsnThrSerGluAspAspGluTyrT 464

 415 ...................AAAATTGTGAAACGGAATAATTATAAAGCA 444
                         :::      ||||||||||:::   ||||:::
 464 hrGlySerGlyAsnSerAlaSerThrProLysArgAsnLysGluLysThr 480

 445 GGGACTAAAGGCCATCCTTATGGCGGCGATTATCATATGCCGCGTTTGCA 494
          |||           :::   :::|||:::          :::   :
 481 .ThrLeuThrAsnThrThrLeuGluSerIleLeuLysLysGlyThrPheV 497

 495 TAAATTTGTCACAGATGCAGAACCTGTTGAAATGACCAGTTATATGGATG 544
     ::|||::::::      |||:::                   |||::::::
 497 alAsnIleThrAlaAsnGlnArg.................IleTyrVal 507

 545 GGCGGAAATATATCGATCAAAATAATTACCCTGACCGTGTTCGTATTGGG 594
     ::::::::|||:::::::::           :::||||||::::::      ||
 508 AsnSerSerIleAsnLeuSerAsnGlySerLeuThrLeuTrpSerGluGl 524

 595 GCAGGCAGGC..........................AATATTGGCGATC 617
     |::::::|||                           :::    ||||||
 524 yArgSerGlyGlyGlyValGluIleAsnAsnAspIleThrThrGlyAspA 541

 618 TGATGAAGATGAGCCCAATAACC.................GCGAAAGTT 649
         :::   :::   :::|||                   :::   |||
 541 spThrArgGlyAlaAsnLeuThrIleTyrSerGlyGlyTrpValAspVal 557

 650 CATATCATATTGCAAGTGCGTATTCTTGGCTCGTTGGTGGCAATACCTTT 699
     |||       :::::::::           |||:::::::   :::    |||
 558 HisLysAsnIleSerLeuGlyAlaGlnGlyAsnIle..AsnIleThrAla 573

 700 GCACAAAAT...............GGATCA...............GGTGG 719
        |||:::                  ||||||                |||
 574 LysGlnAspIleAlaPheGluLysGlySerAsnGlnValIleThrGlyGl 590

 720 TGGCACAGTCAACTTAGGTAGTGAAAAAATTAAACATAGCCCATATGGTT 769
       ||||||::::::    |||::::::|||                   ||||
 590 nGlyThrIleThrSerGlyAsnGlnLys.................GlyP 601

 770 TTTTACCAACAGGAGGCTCATTTGGCGACAGTGGCTCACCAATGTTTATC 819
     ||        ::::::    |||::::::    :::||||||    :::    :::
 601 heArgPheAsnAsnValSerLeuAsnGlyThrGlySerGlyLeuGlnPhe 617

 820 TATGATGCCCAAAAGCAAAAGTGGTTAATTAATGGGGTATTGCAAACGGG 869
            :::    :::|||:::    |||::::::    ::::::
 618 ThrThrLysArgThrAsnLysTyrAlaIleThrAsnLysPheGluGlyTh 634

 870 CAACCCCTATATAGGAAAAAGCAAT...GGCTTCCAGCTGGTTCGTAAAG 916
                    ||||||   |||    ::::::    |||    ::::::
 634 rLeuAsnIleSerGlyLysValAsnIleSerMetValLeuProLysAsnG 651

 917 ATTGGTTCTATGATGAAATCTTTGCTGGAGATACCCAT......TCAGTA 960
     ::        ||||||:::   |||    |||    |||:::        :::
 651 luSerGlyTyrAspLys...PheLysGlyArgThrTyrTrpAsnLeuThr 666

 961 TTCTACGAACCACGTCAAAATGGGAAATACTCTTTTAACGACGATAATAA 1010
           :::       ::::::|||::::::::::::::    |||::::::
 667 SerLeuAsnValSerGluSerGlyGluPheAsnLeuThrIleAspSerAr 683

1011 TGGCACA.........GGAAAAATCAATGCCAAACATGAACACAATTCTC 1051
     :|||:::              |||    ::::::     ::::::  |||:::
 683 gGlySerAspSerAlaGlyThrLeuThrGlnProTyrAsnLeuAsnGlyI 700

1052 TGCCT...AATAGATTAAAAACACGAACCGTTCAATTGTTTAATGTTTCT 1098
     ::       |||:::      |||    :::|||:::       |||:::
 700 leSerPheAsnLysAspThrThrPheAsnValGluArg...AsnAlaArg 715

1099 TTATCCGAGACAGCAAGAGAACCTGTTTATCATGCTGCAGGTGGTGTCAA 1148
     ::::::             :::   |||:::                |||:::||

 716 ValAsnPheAspIleLysAlaProIle..............GlyIleAs 727

1149 CAGTTATCGACCCAGACTGAATAATGGAGAAAATATTTCCTTTATTGACG 1198
     |:::|||                        ::::::::::::::   :::
 727 nLysTyr....................SerSerLeuAsnTyrAlaSer. 736

1199 AAGGAAAAGGCGAATTGATACTTACCAGCAACATCAATCAAGGTGCTGGA 1248
                        :::::::::|||||||:::    ::::::|||
 737 ...................PheAsnGlyAsnIleSerValSerGlyGly 746

1249 GGATTATATTTCCAAGGAGATTTTACGGTCTCGCCTGAAAATAACGAAAC 1298
     |||          :::   ||||||||||:::      :::::::::::::
 747 Gly.........SerValAspPheThrLeuLeuAlaSerSerSerAsnVa 760

1299 TTGGCAAGGCGCGGGCGTT....................CATATCAGTG 1327
     :    |||       ||||||                      ::::::|||
 760 l...GlnThrProGlyValValIleAsnSerLysTyrPheAsnValSerT 776

1328 AAGACAGTACCGTTACTTGGAAAGTAAACGGC.................. 1359
           |||::::::    :::|||::::::|||
 776 hrGlySerSerLeuArgPheLysThrSerGlySerThrLysThrGlyPhe 792

1360 ...GTGGCAAACGAC...CGCCTGTCCAAAATCGGCAAAGGCACGCTGCA 1403
        :::   :::|||     |||:::     |||   :::
 793 SerIleGluLysAspLeuThrLeuAsnAlaThrGlyGlyAsnIleThrLe 809

1404 CGTTCAAGCCAAAGGGGAAAAC..........................C 1426
      :::|||::::::|||   :::                            :
 809 uLeuGlnValGluGlyThrAspGlyMetIleGlyLysGlyIleValAlaL 826

1427 AAGGCTCGATCAGCGTGGGCGACGGTACAGTCATTTTGGATCAGCAGGCA 1476
     ::    :::|||:::          |||::::::    :::
 826 ysLysAsnIleThrPheGluGlyGlyAsnIleThrPhe............ 838

1477 GACGATAAAGGCAAAAAACAAGCCTTTAGTGAAATCGGCTTGGTCAGCGG 1526
                |||::::::::::|||    :::||||||        :::||
 839 .........GlySerArgLysAlaValThrGluIle.........GluGl 849

1527 CAGGGGTACGGTGCAACTGAATGCCGATAATCAGTTCAACCCCGACAAAC 1576
     |:::    |||::::::    ||||||:::
 849 yAsnValThrIleAsnAsnAsnAlaAsn...................... 858

1577 TCTATTTCGGCTTTCGCGGCGGACGTTTGGATTTAAACGGGCATTCGCTT 1626
                                   :::    |||    |||    ::::::
 859 ........................ValThrLeuIleGlySerAspPhe 866

1627 TCGTTCCACCGTATTCAAAATACCGATGAAGGGGCGATGATTGTCAACCA 1676
     :::    |||:::          |||    :::       :::|||:::|||
 867 AspAsnHisGlnLysProLeuThrIleLysLysAspValIleIleAsnSe 883

1677 CAATCAA......GACAAAGAATCCACCGTTACCATTACAGGCAATAAAG 1720
     ::::::                 :::    |||:::|||:::||||||
 883 rGlyAsnLeuThrAlaGlyGlyAsnIleValAsnIleAlaGlyAsnLeuT 900

1721 ATATTGCTACAACCGGCAAT.............................. 1740
        :::    :::::::::|||
 900 hrValGluSerAsnAlaAsnPheLysAlaIleThrAsnPheThrPheAsn 916

1741 ...............AACAACAGCTTGGATAGCAAAAAAGAAATTGCCTA 1775
                   :::|||:::   :::|||:::   :::||||||
 917 ValGlyGlyLeuPheAspAsnLysGlyAsnSerAsnIleSerIleAlaLy 933

1776 CAACGGT...TGGTTTGGCGAGAAAGATACGACCAAAACGAACGGGCGGC 1822
      :::|||       |||    :::    |||::::::|||          :::|
 933 sGlyGlyAlaArgPheLysAspIleAspAsnSerLys.........AsnL 947

1823 TCAACCTTGTTTACCAGCCCGCCGCAGAAGACCGCACCCTGCTGCTTTCC 1872
     ||:::::::::   :::   ::::::          ||||||::::::   :::
 947 euSerIleThrThrAsnSerSerSerThrTyrArgThrIleIleSerGly 963

1873 GGCGGAACAAATTTAAACGGCAACATCACGCAAACAAACGGCAAACTGTT 1922
     :::    ||||||   ||||||:::::::::      ||||||
 964 AsnIleThrAsnLysAsnGlyAspLeuAsnIleThrAsn........... 976
```

```
1923 TTTCAGCGGCAGACCAACACCGCACGCCTACAATCATTTAAACGACCATT 1972
       :::|||       |||                :::::::::
 977 ....GluGlySerAspThrGlu.........MetGlnIleGlyGlyAspV 989

1973 GGTCGCAAAAAGAGGGCATTCCTCGCGGGGAAATCGTGTGGGACAACGAC 2022
       |||||||||||             |||::::::::  ::::::|||
 989 alSerGlnLysGlu............GlyAsnLeuThrIleSerSerAsp 1001

2023 TGGATCAACCGCACATTTAAAGCGGAAAACTTCCAAATTAAAGGCGGACA 2072
       ||||||  |||             :::::::::  |||||||::||| 
1002 LysIleAsnIleThr.........LysGlnIleThrIleLysAlaGly.. 1014

2073 GGCGGTGGTTTCCCGCAATGTTGCCAAAGTGAAAGGCGATTGGCATTTGA 2122
                         |||       ::::::|||        :
1015 ..................ValAspGlyGluAsnSerAspSerAspAlaT 1025

2123 GCAATCACGCCCAAGCAGTTTTGGTGTCGCACCGCATCAAAGCCACACA 2172
     ::|||:::|||:::                                |||
1025 hrAsnAsnAlaAsnLeu.............................Thr 1031

2173 ATCTGTACACGTTCGGACTGGACGGGTCTGACAAATTGTGTCGAAAAAAC 2222
     |||   |||:::                             |||
1032 IleLysThrLys.............................GluLeuLy 1038

2223 CATTACCGACGATAAAGTGATTGCTTCATTGACTAAGACCGACATCAGCG 2272
      :::|||:::|||      |||:::::::::::|||::::::|||::::
1038 sLeuThrGlnAspLeuAsnIleSerGlyPheAsnLysAlaGluIleThrA 1055

2273 GCAATGTCGATCTTGCCGATCACGCTCATTTAAATCTCACAGGGCTTGCC 2322
     :::::   |||   :::|||
1055 laLys...AspGlySerAsp..........................Leu 1061

2323 ACACTCAACGGCAATCTTAGTGCAAATGGCGATACACGTTATACAGTCAG 2372
     |||::::::::::::   ||||||::|||    :::        |||::
1062 ThrIleGlyAsnThrAsnSerAlaAspGlyThrAsnAlaLysLysValTh 1078

2373 CCACAACGCCACCCAAAACGGCAACCTTAGCCTCGTGGGCAATGCCCAAG 2422
     :   |||    :::::::::::::::::::|||      |||
1078 rPheAsnGlnValLysAspSerLysIleSerAlaAspGly.......... 1091

2423 CAACATTTAATCAAGCCACATTAAACGGCAACACATCGGCTTCGGGCAAT 2472
             :::::::::||||||::::::::::::::::::|||||||:::
1092 ........HisLysValThrLeuHisSerLysValGluThrSerGlySer 1105

2473 GCTTCATTTAATCTAAGCGACCACGCCGTACAAAACGGCAGTCTGACGCT 2522
         :::   |||   :::|||   :::    :::|||::::::||||||::
1106 AsnAsn...AsnThrGluAspSerSerAspAsnAsnAlaGlyLeuThrIl 1121

2523 TTCCGGCAACGCTAAGGCAAACGTAAGCCATTCCGCACTCAACGGTAATG 2572
     :      :::||||||   ||||||:::         :::|||:::|||:
1121 e......AspAlaLys...AsnValThr.........ValAsnAsnAsnI 1132

2573 TCTCCCTAGCCGATAAGGCAGTATTCCATTTGAAAGCAGCCGCTTTACC 2622
     :::::   :::   |||||||||         :::   |||    :::
1132 leThr...SerHisLysAlaVal.........SerIleSerAlaThrSer 1144

2623 GGACAAATCAGCGGCGGCAAGGATACGGCATTACACTTAAAA......GA 2666
     |||:::|||:::             |||:::::::::          ::
1145 GlyGluIleThrThrLysThrGlyThrThrIleAsnAlaThrThrGlyAs 1161

2667 CAGCGAATGGACGCTGCCGTCAGGCACGGAATTAGGCAATTTAAACCTTG 2716
     :   |||   |||       :::|||:::   ||||||:::::::::   :
1161 nValGluIleThrAlaGlnThrGlySerIleLeuGlyGlyIleGluSerS 1178

2717 ACAACGCCACCATTACACTCAATTCCGCCTATCGCCACGATGCGGCAGGG 2766
     :::::::::::::::||||||:::::::::             |||::::::
1178 erSerGlySerValThrLeuThrAlaThrGluGlyAlaLeuAlaValSer 1194

2767 GCGCAAACCGGCAGTGCGACAGATGCGCCGCGCCGCCGTTCGCGCCGTTC 2816
           :::|||:::::::::   :::      :::   :::      ::
1195 AsnIleSerGlyAsnThrValThrValThrAlaAsnSerGlyAlaLeuTh 1211

2817 GCGCCGTTCCCTATTATCCGTTACACCGCCAACTTCGGTAGAATCCCGTT 2866
     :      :::      ::::::            ||||||   :::
1211 rThrLeuAlaGlySerThrIleLysGlyThrGluSerValThrThrSerS 1228

2867 TCAACACGCTGACGGTAAACGGCAAATTGAACGGTCAGGGAACA...... 2910
       ::::::      :::::|||   ::::::|||   ||||||
1228 erGlnSerGlyAspIleGlyGlyThrIleSerGly...GlyThrValGlu 1243

2911 TTCCGCTTTATGTCGGAACTCTTCGGCTACCGCAGCGACAAATTGAAGCT 2960
       :::      ::::::|||            ::::::|||:::|||
1244 ValLysAlaThrGluSerLeuThrThrGlnSerAsnSerLysIleLys.. 1259

2961 GGCGGAAAGTTCCGAAGGCACTTACACCTTGGCGGTC............ 2997
      |||   ::::::|||::::::   |||   |||:::
1260 .AlaThrThrGlyGluAlaAsnValThrSerAlaThrGlyThrIleGlyG 1276

2998 ..........AACAATACCGGCAACGAACCTGCAAGCCTCGAACAATTG 3036
                :::||||||   |||      |||:::      :::|||
1276 lyThrIleSerGlyAsnThrValAsnValThrAlaAsnAlaGlyAspLeu 1292

3037 ACGGTA........................GTGGAAGGAAAAGACAACAA 3062
     ||||||                        :::||||||      :::
1293 ThrValGlyAsnGlyAlaGluIleAsnAlaThrGluGlyAlaAlaThrLe 1309

3063 ACCGCTGTCCGAAAACCTTAATTTCACCCTGCAAAACGAACACGTCGATG 3112
              |||::::::         |||      ::::::|||:::   :
1309 uThrThrSerSerGlyLysLeuThrThrGluAlaSerSerHisIleThrS 1326

3113 CCGGCGCGTGGCGTTACCAACTCATCCGCAAAGACGGCGAGTTCCGCCTG 3162
      :: |||:::|||   :::|||   ||||||:::   :::|||   |||
1326 er.AlaLysGlyGlnValAsnLeuSerAlaGlnAspGlySerValAlaGl 1342

3163 CATAATCCGGTCAAAGAACAAGAGCTTTCCGACAAACTCGGCAAGGCAGA 3212
        |||::::::   |||   ::::::   |||:::::::::
1342 ySerIleAsnAlaAlaAsnValThrLeuAsnThrThrGlyThrLeuThrT 1359

3213 AGCCAAAAAACAGGCGGAAAAAGACAACGCGCAAAGCCTTG 3253
        |||:::   :::   ::::::|||      :::|||
1359 hrValLysGlySerAsnIleAsnAlaThrSerGlyThrLeu 1372
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01841

```
seq_documentation_block:
ID   AAB01841 standard; Protein; 1004 AA.
XX
AC   AAB01841;
XX
DT   11-SEP-2000  (first entry)
XX
DE   Haemophilus influenzae strain PMH1 mature HMW2A protein, SEQ ID NO:57.
XX
KW   Mature HMW protein; hmw gene; hmwA1; hmwA2; high molecular weight;
KW   non-typeable Haemophilus influenzae; NTHi; non-encapsulated;
KW   recombinant production; Escherichia coli; antibacterial; vaccine;
KW   human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;
KW   detection; diagnosis.
XX
OS   Haemophilus influenzae strain PMH1.
XX
PN   WO200020609-A2.
XX
PD   13-APR-2000.
XX
PF   07-OCT-1999;   99WO-CA00938.
XX
PR   07-OCT-1998;   98US-0167568.
PR   08-DEC-1998;   98US-0206942.
XX
PA   (CONN-) CONNAUGHT LAB LTD.
XX
PI   Loosmore SM,  Yang Y,  Klein MH;
XX
DR   WPI; 2000-303789/26.
```

```
DR   N-PSDB; AAA52190.
XX
PT   Nucleic acid molecule for producing recombinant high molecular weight
PT   proteins of Haemophilus which are used as a vaccine to provide
PT   protection against Haemophilus induced diseases in humans -
XX
PS   Claim 8; Fig 25A-O; 307pp; English.
XX
CC   The invention relates to the recombinant production of Haemophilus
CC   influenzae high molecular weight (HMW) proteins in Escherichia coli. The
CC   expression construct used to effect recombinant expression comprises a
CC   promoter functional in E. coli (e.g., the T7 promoter) operably linked
CC   to a modified hmwABC operon from a non-typeable (non-encapsulated) H.
CC   influenzae (NTHi). Most HMW-expressing NTHi strains contain two hmw gene
CC   clusters termed hmw1ABC and hmw2ABC. Each hmwABC operon comprises hmwA,
CC   hmwB and hmwC genes. The hmwA genes encode the structural HMWA proteins
CC   and the hmwB and hmwC genes encode accessory proteins which are
CC   responsible for post-translational processing and secretion of the HMWA
CC   proteins. The modified hmwABC operon used in the expression construct of
CC   the invention contains an A gene modified such that it encodes only the
CC   mature HMWA. The invention also discloses hmwA genes (AAA52175-A52198)
CC   and HMWA proteins (AAB01824-B01849) from the non-typeable H. influenzae
CC   strains Joyc, K1, K21, LCDC2, PMH1, 15 and 12. The nucleic acids and
CC   vectors are used for the production of recombinant H. influenzae HMW
CC   proteins which can be used as vaccines to mediate a humoral or
CC   cell-mediated immune response to provide protection against diseases in
CC   humans caused by H. influenzae (e.g., otitis media, epiglottitis,
CC   pneumonia and tracheobronchitis). The HMW proteins are also useful as
CC   antigens in immunoassays for detecting antibodies against Haemophilus,
CC   HMW proteins and/or HMW peptides. The nucleotide sequences encoding the
CC   HMW proteins can be used to isolate and clone hmw genes from other
CC   non-typeable strains of Haemophilus via hybridisation reactions. The
CC   present sequence represents a mature HMWA protein from a non-typeable
CC   strain of H. influenzae.
XX
SQ   Sequence   1004 AA;


alignment_scores:
          Quality:  274.00          Length:   1077
            Ratio:  0.531             Gaps:     49
 Percent Similarity:  47.911   Percent Identity:  20.149

alignment_block:
 US-09-303-518D-649 x AAB01841    ..

 Align seg 1/1  to: AAB01841  from: 1  to: 1004

    634 AATAACCGCGAAAGTTCATATCATATT....................... 660
        ||||||:::::||||||       |||
     86 AsnAsnLysLysSerSerValLysIleAsnGlyAsnIleThrSerThrTh 102

    661 ..........GCAAGTGCGTATTCT......TGGCTC............. 681
                  :::   ||||||       |||:::
    102 rAsnGlyAsnLeuThrIleTyrSerSerGlyTrpValAspIleHisLysA 119

    682 ..GTTGGTGGCAATACCTTTGCACAAAATGGATCAGGTGGTGGCACAGTC 729
          :::       ||||||         |||    ::::::|||||||:::|||
    119 snIleThrLeuAsnThrGlyTyrLeuAsnIleThrAlaGlyGlySerVal 135

    730 .........AACTTAGGTAGTGAAAAAATTAAACATAGCCCATATGGTTT 770
                 :::   |||:::||||||   ::::::          :::
    136 AlaPheGluLysAlaGlyAsnGluLysGlyArgGlnValSerGluSerVa 152

    771 TTTACCAACAGGAGGCTCATTTGGCGACAGTGGCTCACCAATGTTTATCT 820
          :::   :::   |||   :::   ::::::   :::      |||    :
    152 lIleLysAlaGlnGlyValIleThrSerGlyValGlyGluGlyPheArgP 169

    821 ATGATGCCCAAAAGCAAAAGTGGTTAATTAATGGGGTATTGCAAACGGGC 870
        :::::       :::              :::|||||||||      :::|||
    169 heAsnAsnValSer............LeuAsnGlyVal...GlyAlaGly 180

    871 AACCCCTATATAGGAAAAAGCAATGGCTTCCAGCTGGTTCGTAAAGATTG 920

            ::::::|||::::::|||         :::       ::::::||
    181 LeuArgPheValGlyGlnLysAsn.........IleSerSerAsnSerTr 194

    921 GTTCTATGATGAAATCTTTGCTGGAGATACCCATTCAGTATTCTACGAAC 970
        |
    194 p................................................. 194

    971 CACGTCAAAATGGG...AAATACTCTTTTAACGACGATAATAATGGCACA 1017
           |||:::|||      |||      |||:::   :::    |||   :::
    195 ..ArgGluAsnThrIleLysAsnArgPheAspGlyAsnLeuAsnIleSer 210

   1018 GGAAAAATCAATGCCAAACATGAA.................CACAATTC 1049
        ||||||:::|||::::::   :::                   |||:::
    211 GlyLysValAsnValSerMetAspValSerGlyThrLysTrpHisThrAr 227

   1050 TCTGCCTAATAGA....................TTAAAAACACGAACCG 1078
          :::    :::|||                  |||::::::    :::
    227 gIleAsnGlyArgThrTyrTrpAsnValThrThrLeuAsnValAlaSerG 244

   1079 TTCAATTGTTTAATGTTTCTTTATCCGAGACAGCAAGAGAACCTGTTTAT 1128
           :::   ||||||:::|||::::::
    244 lySerSerPheAsnLeuSerIleAsp....................... 252

   1129 CATGCTGCAGGTGGTGTCAACAGTTATCGACCCAGACTGAATAATGGAGA 1178
                |||:::|||::::::|||              :::|||    ::
    253 ......AlaSerGlyIleSerSer..............GlyAsnGlnAs 262

   1179 AAATATTTCCTTTATTGACGAAGGAAAAGGCGAATTGATACTTACCAGCA 1228
        :::|||:::
    262 pAspIleThr....................................... 265

   1229 ACATCAATCAAGGTGCTGGAGGATTATATTTCCAAGGAGATTTTACGGTC 1278
               |||:::|||   :::|||:::   |||:::|||:::    |||
    266 .....AsnArgGlyLeuAsnGlyIleThrPheAsnGlyGluAsnThrPhe 280

   1279 TCGCCTGAAAATAACGAAACTTGGCAAGGCGCGGCGTTCATATCAGTGA 1328
        :::       :::::::::|||          |||:::    ||||||    ::
    281 AsnIleAlaGlnGlySerThr.........AlaAsnPheHisIle...Ly 293

   1329 AGACAGTACCGTTACTTGGAAAGTAAAC.........GGCGTGGCAAACG 1369
        :    |||:::::::|||    |||    |||          ::::::    |||
    293 sThrSerValMetThrProLysProAsnSerAsnTyrAlaLeuPheAsnG 310

   1370 ACCGCCTGTCCAAAATCGGCAAAGGCACGCTGCACGTTCAAGCCAAAGGG 1419
          ::::::|||    :::|||    ||||||:::::    :::    ::::::
    310 lyAsnIleSerValLeuGlyGlyGlyThrValAsnPheGluLeuAsnAla 326

   1420 GAAAACCAAGGCTCGATCAGCGTGGGCGACGGTACAGTCATTTTGGATCA 1469
        ::::::       |||      :::    :::|||::::::|||    :::||
    327 SerSer......SerThrHisThrThrSerGlyAlaIleIleAsnSerGl 341

   1470 GCAGGCAGACGATAAAGGCAAAAAACAAGCCTTTAGTGAAATCGGCTTGG 1519
        |:::    :::                                          |
    341 nAsnPheAsn.......................................V 345

   1520 TCAGCGGCAGGGGTACGGTGCAACTGAATGCCGATAATCAGTTCAAC... 1566
        ||||||||    :::     ::::::|||:::|||:::::::::   |||
    345 alSerGlyGlySerLysLeuAsnLeuLysAlaSerGlySerThrAsnThr 361

   1567 ............CCCGACAAACTCTATTTCGGCTTTCGCGGCGGACGTTT 1604
                    ::::::|||    ::::::       ||||||::::
    362 AlaPheLeuIleLysAsnAsnLeuThrLeuAsnAlaThrGlyGlyAsnIl 378

   1605 GGATTTAAACGGGCATTCGCTTTCGTTCCACCGTATTCAA.......... 1644
        :::::::::      :::    :::       |||||||||
    378 eGluIleLysGlnValGluGlyThrAspSerArgIleGlnLysGlyValV 395

   1644 .................................................. 1644

    395 alAlaGluGlnAsnIleIlePheGluGlyGlyAsnIleThrLeuGlySer 411

   1644 .................................................. 1644
```

```
 412 GlnLysAlaProThrGluIleLysGlyAspValThrValLysGlnGlyTh 428

1645 ...........................AATACCGATGAAGGGGCGATGA 1666
                              ::::::   :::||||||:::|
 428 rAsnAlaThrLeuArgSerAlaAsnPheAspAsnHisLysGlyAlaLeuI 445

1667 TTGTCAACCACAATCAAGACAAAGAA....................TCC 1695
     |||||||   |||           :::                   :::
 445 leValAsnGlyAsnValThrAlaAsnGlyAsnLeuThrAlaAspGlyAsp 461

1696 ACCGTTACCATTACAGGCAATAAAGATATTGCTACAACCGGCAATAACAA 1745
     |||:::   |||   ||||||   |||:::|||      ::::::   ||
 462 ThrIleLysIleLysGlyAsnLeuAspValAlaGlnGlyAlaLysPheAs 478

1746 CAGCTTGGATAGCAAAAAAGAAATTGCCTACAACGGTTGGTTTGGCGAGA 1795
     |:::   ::::::|||:::::::::         :::|||   |||   :::
 478 nGly...SerThrLysAsnAsnLeuAsnIleThrGlyThrPheThrAsnA 494

1796 AAGATACGACC.........AAAACGAACGGGCGGCTCAACCTTGTTTAC 1836
     ::   |||:::                 |||:::|||   :::||||||
 494 snGlyThrSerIleIleAspIleThrGlnGlyValValAsnLeu...... 508

1837 CAGCCCGCCGCAGAAGACCGCACCCTGCTGCTTTCCGGCGGAACAAATTT 1886

 508 .................................................. 508

1887 AAACGGCAACATCACGCAAACAAACGGCAAACTGTTTTTCAGCGGCAGAC 1936
         ||||||:::|||    ::::::|||||||||           :::
 509 ....GlyAsnValThr...AsnAspGlyLysLeu...........AsnI 519

1937 CAACACCGCACGCCTACAATCATTTAAACGACCATTGGTCGCAAAAAGAG 1986
        |||   ||||||                          |||   ::::::
 519 leThrThrHisAlaLys....................SerGlyGlnLys 528

1987 GGCATTCCTCGCGGGGAAATCGTGTGGGACAACGACTGGATCAACCGCAC 2036
     :::|||   ||||||:::|||:::        :::      :::|||   ||
 529 SerIleIleArgGlyAspIleIleAsnLysGlnGlyAsnLeuAsnIleTh 545

2037 ATTTAAAGCGGAAAACTTCCAAATTAAAGGCGGACAGGCGGTGGTTTCCC 2086
     |   :::   :::|||   :::|||:::                  ::::::
 545 rAspAsnAsnSerAsnAlaGluIleGlu..............IleGlyG 557

2087 GCAATGTTGCCAAAGTGAAAGGCGATTGGCATTTGAGCAATCACGCCCAA 2136
        |||:::::::::   :::|||:::      :::|||:::
 557 lyAsnIleSerGlnLysGluGlyAsnLeuThrIleSerSerAspLysVal 573

2137 GCAGTT....................TTTGGTGTCGCACCGCATCAAAG 2165
        :::                             ||||||      :::||
 574 AsnIleThrLysGlnIleThrIleLysAlaGlyVal...AspGlyGluSe 589

2166 CCACACAATCTGTACACGTTCGGACTGGACGGGTCTGACAAATTGTGTCG 2215
     |   :::         |||   ||||||   :::         :::      |
 589 rSerSerSerSerThrAlaSerAspAlaAsnLeuThrIleLysThrLysG 606

2216 AAAAAACCATTACCGACGATAAAGTGATTGCTTCATTGACTAAGACCGAC 2265
     ||   |||:::||||||:::         |||::::::::::::|||::::::
 606 luLeuThrPheThrAspAsnLeuAsnIleSerGlyPheAsnLysAlaGlu 622

2266 ATCAGCGGC......AATGTCGATCTTGCCGATCACGCTCATTTAAATCT 2309
     |||::::::         |||   ||||||
 623 IleThrAlaLysAspAsnSerAspLeu....................... 631

2310 CACAGGGCTTGCCACACTCAACGGCAATCTTAGTGCAAATGGCGATACAC 2359
                       :::   |||:::   |||:::::::::::::::
 632 ...............IleIleGlyLysAlaSerSerAspAsnSerAsnA 643

2360 GTTATACAGTCAGCCACAACGCCACCCAAAACGGCAACCTTAGCCTCGTG 2409
                  |||:::   :::   ::::::::::::::::::|||   :::
 643 laLysGlnValThrPheAspLysValLysAspSerLysIleSer...Ala 658

2410 GGCAATGCCCAAGCAACATTTAATCAAGCCACATTAAACGGCAACACATC 2459
     ||||||   ::::::|||:::|||:::   :::   ::::::|||:::::
 659 GlyAsnHisAsnValThrLeuAsnSerLysValGluThrSerAsnSerAs 675

2460 GGCTTCG...GGCAATGCTTCA............................ 2478
     :::::|||   ||||||:::|||
 675 pGlySerThrGlyAsnGlySerAspAspAsnAsnIleGlyLeuThrIleS 692

2479 ......................TTTAATCTAAGCGACCACGCCGTACAA 2505
                           |||:::::::::|||   :::
 692 erAlaLysAspValThrValAsnSerAsnIleThrSerHisLysThrVal 708

2506 AACGGCAGTCTGACGCTTTCCGGC...AACGCTAAGGCAAACGTAAGCCA 2552
     |||   |||   :::   :::|||   ::::::||||||::::::::::
 709 AsnIleSerAlaSerGluGlyGlyIleThrThrLysAlaGlyThrThrIl 725

2553 TTCCGCACTCAACGGTAATGTCTCCCTAGCCGATAAGGCAGTATTCCATT 2602
     :::|||   :::|||:::|||:::::::::   |||
 725 eAsnAlaThrThrGlySerValGluValThrAlaLys............ 737

2603 TTGAAAGCAGCCGCTTTACCGGACAAATCAGCGGC.........GGCAAG 2643
                   ||||||:::|||||||||            ||||||
 738 .................ThrGlyAspIleSerGlyThrIleSerGlyLys 748

2644 GATACGGCATTACACTTAAAAGACAGCGAATGGACGCTGCCGTCAGGC.. 2691
        :::::::::      :::   ::::::   |||:::   ::::::
 749 ThrValSerValThrAlaSerThrGlyAspLeuThrValArgLysAlaAl 765

2692 .................................ACGGAATTAGGCAATT 2707
                                      |||      ||||||
 765 aThrIleSerValThrGluGlyAlaAlaThrLeuThrAlaThrGlyAsnT 782

2708 TAAACCTTGACAACGCCACCATTACACTCAATTCCGCCTATCGCCACGAT 2757
              :::|||      :::::::::|||:::         :::
 782 hrLeuThrThrGluAlaGlySerSerIleThrSerThrLysGlyGlnVal 798

2758 GCGGCAGGGGCGCAAACCGGCAGT...GCGACAGATGCGCCGCGCCGCCG 2804
        :::||||||   ||||||   |||      :::               ::
 799 AspLeuSerAlaGlnAspGlySerIleAlaGlyGlnIleSerAlaAlaAs 815

2805 TTCGCGCCGTTCGCGCCGTTCCCTATTATCCGTTACACCGCCAACTTCGG 2854
     :            :::      :::   |||:::::::::      :::::::
 815 nValThrLeuAsnThrThrGlyThrLeuThrThrValGluGlySerAsnI 832

2855 TAGAATCCCGTTTCAACACGCTGACGGTAAACGGCAAATTGAACGGTCAG 2904
     ::::::::      :::||||||::::::|||:::|||
 832 leLysAlaThrSerGlyThrLeuAlaIleAsnAlaLys........... 844

2905 GGAACATTCCGCTTTATGTCGGAACTCTTCGGCTACCGCAGCGACAAATT 2954
                                              :::   |||||
 845 ......................................AspAlaLysLe 848

2955 GAAGCTGGCGGAAAGTTCCGAAGGCACTTACACCTTGGCGGTCAACAATA 3004
     |      :::   |||:::::::   |||   :::   |||:::|||   :
 848 uAspGlyThrAlaSerGlyAsnArgThrGluValAsnAlaThrAsnAlaS 865

3005 CCGGCAACGAACCTGCAAGCCTCGAACAATTGACGGTAGTGGAAGGAAAA 3054
     ::|||:::                                 |||:::
 865 erGlySer...................................GlySer 869

3055 GACAACAAACCGCTGTCCGAAAACCTTAATTTCACCCTGCAAAACGAACA 3104
        :::            |||:::|||:::|||:::|||   :::   :::
 870 ValThrAlaLysThrSerSerAsnValAsnIleThrGlyAspLeuSerTh 886

3105 CGTCGATGCCGGCGCGTGGCGTTACCAACTCATCCGCAAAGACGGCGAGT 3154
     :::::::::              ::::::|||   ::::::|||
 886 rIleAsnGly...........LeuAsnIleIleSerGluAsnGly.... 897

3155 TCCGCCTGCATAATCCGGTCAAAGAACAAGAGCTTTCCGACAAACTCGGC 3204
              :::|||   |||                          :::|||
 898 ........ArgAsnThrVal....................ArgLeuArg 904

3205 AAGGCAGAAGCCAAAAAACAGGCGGAAAAAGACAACGCGCAAAGCCTTGA 3254
           |||         :::      :::   ::::::   |||::::::
 905 GlyLysGluIleAspValLysTyrIleGlnProGlyValAlaSerValGl 921
```

```
3255 CGCGCTGATTGCGGCCGGGCGCGATGCCGTCGAAAAGACAGAAAGCGTTG 3304
      :   :::|||   |||   |||   :::::|||||||::::::::::::
 921 uGluValIleGluAlaLysArg...ValLeuGluLysValLysAspLeuS 937

3305 CCGAACCGGCCCGGCAGGCAGGCGGGGAAAATGTCGGCATTATGCAGGCG 3354
     :::::         |||:::::::   ::::::   :::|||:::      :::
 937 erAspGluGluArgGluThrLeuAlaLys...LeuGlyValSerAlaVal 952

3355 GAGGAAGAGAAAAAACGGGTGCAGGCGGATAAAGACACCGCCTTGGCGAA 3404
                  :::   |||:::   ::::::      ::::::
 953 ...........ArgPheValGluProAsnAsnAlaIleThrIleAsnTh 965

3405 ACAGCGCGAAGCGGAAACCCGGCCGGCTACC 3435
       |||:::|||      |||||||||::::::
 965 rGlnAsnGluPheThrThrArgProSerSer 975
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01840

```
seq_documentation_block:
ID   AAB01840 standard; Protein; 1010 AA.
XX
AC   AAB01840;
XX
DT   11-SEP-2000  (first entry)
XX
DE   Haemophilus influenzae strain PMH1 HMW2A protein, SEQ ID NO:55.
XX
KW   HMW protein; hmw gene; hmwA1; hmwA2; high molecular weight;
KW   non-typeable Haemophilus influenzae; NTHi; non-encapsulated;
KW   recombinant production; Escherichia coli; antibacterial; vaccine;
KW   human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;
KW   detection; diagnosis.
XX
OS   Haemophilus influenzae strain PMH1.
XX
PN   WO200020609-A2.
XX
PD   13-APR-2000.
XX
PF   07-OCT-1999;   99WO-CA00938.
XX
PR   07-OCT-1998;   98US-0167568.
PR   08-DEC-1998;   98US-0206942.
XX
PA   (CONN-) CONNAUGHT LAB LTD.
XX
PI   Loosmore SM,  Yang Y,  Klein MH;
XX
DR   WPI; 2000-303789/26.
DR   N-PSDB; AAA52189.
XX
PT   Nucleic acid molecule for producing recombinant high molecular weight
PT   proteins of Haemophilus which are used as a vaccine to provide
PT   protection against Haemophilus induced diseases in humans -
XX
PS   Claim 12; Fig 25A-O; 307pp; English.
XX
CC   The invention relates to the recombinant production of Haemophilus
CC   influenzae high molecular weight (HMW) proteins in Escherichia coli. The
CC   expression construct used to effect recombinant expression comprises a
CC   promoter functional in E. coli (e.g., the T7 promoter) operably linked
CC   to a modified hmwABC operon from a non-typeable (non-encapsulated) H.
CC   influenzae (NTHi). Most HMW-expressing NTHi strains contain two hmw gene
CC   clusters termed hmw1ABC and hmw2ABC. Each hmwABC operon comprises hmwA,
CC   hmwB and hmwC genes. The hmwA genes encode the structural HMWA proteins
CC   and the hmwB and hmwC genes encode accessory proteins which are
CC   responsible for post-translational processing and secretion of the HMWA
CC   proteins. The modified hmwABC operon used in the expression construct of
CC   the invention contains an A gene modified such that it encodes only the
CC   mature HMWA. The invention also discloses hmwA genes (AAA52175-A52198)
CC   and HMWA proteins (AAB01824-B01849) from the non-typeable H. influenzae
CC   strains Joyc, K1, K21, LCDC2, PMH1, 15 and 12. The nucleic acids and
CC   vectors are used for the production of recombinant H. influenzae HMW
CC   proteins which can be used as vaccines to mediate a humoral or
CC   cell-mediated immune response to provide protection against diseases in
CC   humans caused by H. influenzae (e.g., otitis media, epiglottitis,
CC   pneumonia and tracheobronchitis). The HMW proteins are also useful as
CC   antigens in immunoassays for detecting antibodies against Haemophilus,
CC   HMW proteins and/or HMW peptides. The nucleotide sequences encoding the
CC   HMW proteins can be used to isolate and clone hmw genes from other
CC   non-typeable strains of Haemophilus via hybridisation reactions. The
CC   present sequence represents an HMWA protein from a non-typeable strain of
CC   H. influenzae.
XX
SQ   Sequence   1010 AA;
```

```
alignment_scores:
             Quality:  274.00            Length:   1077
               Ratio:  0.531               Gaps:     49
  Percent Similarity:  47.911   Percent Identity: 20.149

alignment_block:
 US-09-303-518D-649 x AAB01840   ..

 Align seg 1/1  to: AAB01840  from: 1  to: 1010

 634 AATAACCGCGAAAGTTCATATCATATT....................... 660
     ||||||::::::||||||      |||
  92 AsnAsnLysLysSerSerValLysIleAsnGlyAsnIleThrSerThrTh 108

 661 .........GCAAGTGCGTATTCT......TGGCTC............ 681
                :::   ||||||       |||:::
 108 rAsnGlyAsnLeuThrIleTyrSerSerGlyTrpValAspIleHisLysA 125

 682 ..GTTGGTGGCAATACCTTTGCACAAAATGGATCAGGTGGTGGCACAGTC 729
       :::      ||||||        |||   ::::::||||||:::|||
 125 snIleThrLeuAsnThrGlyTyrLeuAsnIleThrAlaGlyGlySerVal 141

 730 .........AACTTAGGTAGTGAAAAAATTAAACATAGCCCATATGGTTT 770
              :::   |||:::||||||   ::::::          :::
 142 AlaPheGluLysAlaGlyAsnGluLysGlyArgGlnValSerGluSerVa 158

 771 TTTACCAACAGGAGGCTCATTTGGCGACAGTGGCTCACCAATGTTTATCT 820
      :::   :::   |||   :::   ::::::   :::      |||   :
 158 lIleLysAlaGlnGlyValIleThrSerGlyValGlyGluGlyPheArgP 175

 821 ATGATGCCCAAAAGCAAAAGTGGTTAATTAATGGGGTATTGCAAACGGGC 870
     :::::      :::            :::|||||||||      :::|||
 175 heAsnAsnValSer............LeuAsnGlyVal...GlyAlaGly 186

 871 AACCCCTATATAGGAAAAAGCAATGGCTTCCAGCTGGTTCGTAAAGATTG 920
           ::::::|||::::::|||          :::      ::::::||
 187 LeuArgPheValGlyGlnLysAsn.........IleSerSerAsnSerTr 200

 921 GTTCTATGATGAAATCTTTGCTGGAGATACCCATTCAGTATTCTACGAAC 970
     |
 200 p................................................ 200

 971 CACGTCAAAATGGG...AAATACTCTTTTAACGACGATAATAATGGCACA 1017
       |||:::|||       |||      |||:::   :::   |||   :::
 201 ..ArgGluAsnThrIleLysAsnArgPheAspGlyAsnLeuAsnIleSer 216

1018 GGAAAAATCAATGCCAAACATGAA.................CACAATTC 1049
     ||||||:::|||::::::   :::                  |||:::
 217 GlyLysValAsnValSerMetAspValSerGlyThrLysTrpHisThrAr 233

1050 TCTGCCTAATAGA....................TTAAAAACACGAACCG 1078
      :::   :::|||                    |||:::::::   :::
 233 gIleAsnGlyArgThrTyrTrpAsnValThrThrLeuAsnValAlaSerG 250

1079 TTCAATTGTTTAATGTTTCTTTATCCGAGACAGCAAGAGAACCTGTTTAT 1128
      :::   ||||||:::|||::::::
 250 lySerSerPheAsnLeuSerIleAsp...................... 258

1129 CATGCTGCAGGTGGTGTCAACAGTTATCGACCCAGACTGAATAATGGAGA 1178
```

```
              |||:::|||:::::|||                    :::|||   ::
 259 ......AlaSerGlyIleSerSer..............GlyAsnGlnAs 268

1179 AAATATTTCCTTTATTGACGAAGGAAAAGGCGAATTGATACTTACCAGCA 1228
     ::::|||:::
 268 pAspIleThr....................................... 271

1229 ACATCAATCAAGGTGCTGGAGGATTATATTTCCAAGGAGATTTTACGGTC 1278
           |||:::|||   :::|||:::   |||:::|||:::   |||
 272 .....AsnArgGlyLeuAsnGlyIleThrPheAsnGlyGluAsnThrPhe 286

1279 TCGCCTGAAAATAACGAAACTTGGCAAGGCGCGGGCGTTCATATCAGTGA 1328
     :::     :::::::::|||         |||:::   ||||||   ::
 287 AsnIleAlaGlnGlySerThr.........AlaAsnPheHisIle...Ly 299

1329 AGACAGTACCGTTACTTGGAAAGTAAAC.........GGCGTGGCAAACG 1369
     :  |||::::::|||   |||   |||          ::::::   |||
 299 sThrSerValMetThrProLysProAsnSerAsnTyrAlaLeuPheAsnG 316

1370 ACCGCCTGTCCAAAATCGGCAAAGGCACGCTGCACGTTCAAGCCAAAGGG 1419
        ::::::|||   :::|||   ||||||::::::   :::   ::::::
 316 lyAsnIleSerValLeuGlyGlyGlyThrValAsnPheGluLeuAsnAla 332

1420 GAAAACCAAGGCTCGATCAGCGTGGGCGACGGTACAGTCATTTTGGATCA 1469
     ::::::         |||   :::   :::|||::::::|||   :::||
 333 SerSer......SerThrHisThrThrSerGlyAlaIleIleAsnSerGl 347

1470 GCAGGCAGACGATAAAGGCAAAAAACAAGCCTTTAGTGAAATCGGCTTGG 1519
     |:::   :::                                       |
 347 nAsnPheAsn.......................................V 351

1520 TCAGCGGCAGGGGTACGGTGCAACTGAATGCCGATAATCAGTTCAAC... 1566
     ||||||||   :::   ::::::|||:::|||:::::::::   |||
 351 alSerGlyGlySerLysLeuAsnLeuLysAlaSerGlySerThrAsnThr 367

1567 ...............CCCGACAAACTCTATTTCGGCTTTCGCGGCGGACGTTT 1604
                    :::::|||   ::::::      ||||||:::::
 368 AlaPheLeuIleLysAsnAsnLeuThrLeuAsnAlaThrGlyGlyAsnIl 384

1605 GGATTTAAACGGGCATTCGCTTTCGTTCCACCGTATTCAA.......... 1644
     :::::::::         :::   :::      |||||||||
 384 eGluIleLysGlnValGluGlyThrAspSerArgIleGlnLysGlyValV 401

1644 .................................................. 1644

 401 alAlaGluGlnAsnIleIlePheGluGlyGlyAsnIleThrLeuGlySer 417

1644 .................................................. 1644

 418 GlnLysAlaProThrGluIleLysGlyAspValThrValLysGlnGlyTh 434

1645 ...........................AATACCGATGAAGGGGCGATGA 1666
                                  ::::::   :::||||||:::|
 434 rAsnAlaThrLeuArgSerAlaAsnPheAspAsnHisLysGlyAlaLeuI 451

1667 TTGTCAACCACAATCAAGACAAAGAA....................TCC 1695
     ||||||||   |||               :::                :::
 451 leValAsnGlyAsnValThrAlaAsnGlyAsnLeuThrAlaAspGlyAsp 467

1696 ACCGTTACCATTACAGGCAATAAAGATATTGCTACAACCGGCAATAACAA 1745
     |||:::   |||   ||||||   |||:::|||      ::::::   ||
 468 ThrIleLysIleLysGlyAsnLeuAspValAlaGlnGlyAlaLysPheAs 484

1746 CAGCTTGGATAGCAAAAAAGAAATTGCCTACAACGGTTGGTTTGGCGAGA 1795
     |:::   ::::::|||:::::::::      :::|||   |||   ::::
 484 nGly...SerThrLysAsnAsnLeuAsnIleThrGlyThrPheThrAsnA 500

1796 AAGATACGACC.........AAAACGAACGGGCGGCTCAACCTTGTTTAC 1836
     ::   |||:::              |||:::|||   :::||||||
 500 snGlyThrSerIleIleAspIleThrGlnGlyValValAsnLeu...... 514

1837 CAGCCCGCCGCAGAAGACCGCACCCTGCTGCTTTCCGGCGGAACAAATTT 1886

 514 .................................................. 514

1887 AAACGGCAACATCACGCAAACAAACGGCAAACTGTTTTTCAGCGGCAGAC 1936
        ||||||:::|||   ::::::|||||||||            :::
 515 ......GlyAsnValThr...AsnAspGlyLysLeu...........AsnI 525

1937 CAACACCGCACGCCTACAATCATTTAAACGACCATTGGTCGCAAAAAGAG 1986
       |||   ||||||                          |||   ::::::
 525 leThrThrHisAlaLys.....................SerGlyGlnLys 534

1987 GGCATTCCTCGCGGGGAAATCGTGTGGGACAACGACTGGATCAACCGCAC 2036
     :::|||   ||||||:::|||:::      :::        :::|||   ||
 535 SerIleIleArgGlyAspIleIleAsnLysGlnGlyAsnLeuAsnIleTh 551

2037 ATTTAAAGCGGAAAACTTCCAAATTAAAGGCGGACAGGCGGTGGTTTCCC 2086
     |   :::   :::|||   :::|||:::                  ::::::
 551 rAspAsnAsnSerAsnAlaGluIleGlu..............IleGlyG 563

2087 GCAATGTTGCCAAAGTGAAAGGCGATTGGCATTTGAGCAATCACGCCCAA 2136
       |||:::::::::   :::|||:::      :::|||:::
 563 lyAsnIleSerGlnLysGluGlyAsnLeuThrIleSerSerAspLysVal 579

2137 GCAGTT.....................TTTGGTGTCGCACCGCATCAAAG 2165
        :::                        ||||||          :::||
 580 AsnIleThrLysGlnIleThrIleLysAlaGlyVal...AspGlyGluSe 595

2166 CCACACAATCTGTACACGTTCGGACTGGACGGGTCTGACAAATTGTGTCG 2215
     |   :::      |||   ||||||   :::            :::      |
 595 rSerSerSerSerThrAlaSerAspAlaAsnLeuThrIleLysThrLysG 612

2216 AAAAAACCATTACCGACGATAAAGTGATTGCTTCATTGACTAAGACCGAC 2265
     ||    |||:::|||||||:::        |||::::::::::::|||::::::
 612 luLeuThrPheThrAspAsnLeuAsnIleSerGlyPheAsnLysAlaGlu 628

2266 ATCAGCGGC......AATGTCGATCTTGCCGATCACGCTCATTTAAATCT 2309
     |||::::::         |||   ||||||
 629 IleThrAlaLysAspAsnSerAspLeu....................... 637

2310 CACAGGGCTTGCCACACTCAACGGCAATCTTAGTGCAAATGGCGATACAC 2359
                         :::   |||:::   |||:::::::::::::::
 638 ................IleIleGlyLysAlaSerSerAspAsnSerAsnA 649

2360 GTTATACAGTCAGCCACAACGCCACCCAAAACGGCAACCTTAGCCTCGTG 2409
             |||:::   :::   :::::::::::::::::|||   :::
 649 laLysGlnValThrPheAspLysValLysAspSerLysIleSer...Ala 664

2410 GGCAATGCCCAAGCAACATTTAATCAAGCCACATTAAACGGCAACACATC 2459
     ||||||   ::::::|||:::|||:::   :::   ::::::|||::::::
 665 GlyAsnHisAsnValThrLeuAsnSerLysValGluThrSerAsnSerAs 681

2460 GGCTTCG...GGCAATGCTTCA............................ 2478
     ::::|||   |||||||:::|||
 681 pGlySerThrGlyAsnGlySerAspAspAsnAsnIleGlyLeuThrIleS 698

2479 .......................TTTAATCTAAGCGACCACGCCGTACAA 2505
                              |||:::::::::|||   :::
 698 erAlaLysAspValThrValAsnSerAsnIleThrSerHisLysThrVal 714

2506 AACGGCAGTCTGACGCTTTCCGGC...AACGCTAAGGCAAACGTAAGCCA 2552
     |||   |||   :::   :::|||   ::::::||||||:::::::::
 715 AsnIleSerAlaSerGluGlyGlyIleThrThrLysAlaGlyThrThrIl 731

2553 TTCCGCACTCAACGGTAATGTCTCCCTAGCCGATAAGGCAGTATTCCATT 2602
       :::|||   :::|||:::|||:::::::::   |||
 731 eAsnAlaThrThrGlySerValGluValThrAlaLys............ 743

2603 TTGAAAGCAGCCGCTTTACCGGACAAATCAGCGGC.........GGCAAG 2643
                        ||||||:::|||||||||         ||||||
 744 .................ThrGlyAspIleSerGlyThrIleSerGlyLys 754

2644 GATACGGCATTACACTTAAAAGACAGCGAATGGACGCTGCCGTCAGGC.. 2691
        :::::::::         :::   ::::::   |||:::   ::::::
 755 ThrValSerValThrAlaSerThrGlyAspLeuThrValArgLysAlaAl 771
```

```
2692 .................................ACGGAATTAGGCAATT 2707
                                  |||      ||||||
 771 aThrIleSerValThrGluGlyAlaAlaThrLeuThrAlaThrGlyAsnT 788

2708 TAAACCTTGACAACGCCACCATTACACTCAATTCCGCCTATCGCCACGAT 2757
                :::|||     :::::::::|||:::       :::
 788 hrLeuThrThrGluAlaGlySerSerIleThrSerThrLysGlyGlnVal 804

2758 GCGGCAGGGGCGCAAACCGGCAGT...GCGACAGATGCGCCGCGCCGCCG 2804
              :::||||||  ||||||   |||   :::           ::
 805 AspLeuSerAlaGlnAspGlySerIleAlaGlyGlnIleSerAlaAlaAs 821

2805 TTCGCGCCGTTCGCGCCGTTCCCTATTATCCGTTACACCGCCAACTTCGG 2854
     :        :::      :::     |||:::::::::      :::::::
 821 nValThrLeuAsnThrThrGlyThrLeuThrThrValGluGlySerAsnI 838

2855 TAGAATCCCGTTTCAACACGCTGACGGTAAACGGCAAATTGAACGGTCAG 2904
     ::::::::        :::||||||::::::|||:::|||
 838 leLysAlaThrSerGlyThrLeuAlaIleAsnAlaLys............ 850

2905 GGAACATTCCGCTTTATGTCGGAACTCTTCGGCTACCGCAGCGACAAATT 2954
                                             :::   |||||
 851 .....................................AspAlaLysLe 854

2955 GAAGCTGGCGGAAAGTTCCGAAGGCACTTACACCTTGGCGGTCAACAATA 3004
     |       :::    |||::::::   |||    :::    |||:::|||    :
 854 uAspGlyThrAlaSerGlyAsnArgThrGluValAsnAlaThrAsnAlaS 871

3005 CCGGCAACGAACCTGCAAGCCTCGAACAATTGACGGTAGTGGAAGGAAAA 3054
     ::|||:::                                  |||:::
 871 erGlySer.................................GlySer 875

3055 GACAACAAACCGCTGTCCGAAAACCTTAATTTCACCCTGCAAAACGAACA 3104
       :::          |||:::|||:::|||:::|||    :::    :::
 876 ValThrAlaLysThrSerSerAsnValAsnIleThrGlyAspLeuSerTh 892

3105 CGTCGATGCCGGCGCGTGGCGTTACCAACTCATCCGCAAAGACGGCGAGT 3154
       :::::::::                :::::|||   ::::::|||
 892 rIleAsnGly...........LeuAsnIleIleSerGluAsnGly.... 903

3155 TCCGCCTGCATAATCCGGTCAAAGAACAAGAGCTTTCCGACAAACTCGGC 3204
             :::|||   |||                    :::|||
 904 ........ArgAsnThrVal....................ArgLeuArg 910

3205 AAGGCAGAAGCCAAAAAACAGGCGGAAAAAGACAACGCGCAAAGCCTTGA 3254
           |||          :::    :::    ::::::   |||:::::
 911 GlyLysGluIleAspValLysTyrIleGlnProGlyValAlaSerValGl 927

3255 CGCGCTGATTGCGGCCGGGCGCGATGCCGTCGAAAAGACAGAAAGCGTTG 3304
     :    :::|||    |||    |||    :::::|||||||::::::::::::::
 927 uGluValIleGluAlaLysArg...ValLeuGluLysValLysAspLeuS 943

3305 CCGAACCGGCCCGGCAGGCAGGCGGGGAAAATGTCGGCATTATGCAGGCG 3354
     :::::        |||:::::::   ::::::    :::|||:::       :::
 943 erAspGluGluArgGluThrLeuAlaLys...LeuGlyValSerAlaVal 958

3355 GAGGAAGAGAAAAAACGGGTGCAGGCGGATAAAGACACCGCCTTGGCGAA 3404
                  :::    |||:::    ::::::    ::::::
 959 ...........ArgPheValGluProAsnAsnAlaIleThrIleAsnTh 971

3405 ACAGCGCGAAGCGGAAACCCGGCCGGCTACC 3435
      |||:::|||        ||||||||::::::
 971 rGlnAsnGluPheThrThrArgProSerSer 981
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:AAR41723

seq_documentation_block:

```
ID   AAR41723 standard; Protein; 1536 AA.
XX
AC   AAR41723;
XX
DT   26-APR-1994  (first entry)
XX
DE   High molecular weight protein 1 (HMW1).
XX
KW   HMW; high molecular weight protein; virus; vaccine; influenza;
KW   epitope; immunity; haemophilus influenzae.
XX
OS   Haemophilus influenzae.
XX
PN   WO9319090-A.
XX
PD   30-SEP-1993.
XX
PF   16-MAR-1993;   93WO-US02166.
XX
PR   16-MAR-1992;   92GB-0005704.
XX
PA   (BARE/) BARENKAMP S J.
PA   (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
XX
PI   Barenkamp SJ;
XX
DR   WPI; 1993-320683/40.
DR   N-PSDB; AAQ49506.
XX
PT   High molecular weight surface proteins - of non-typeable
PT   haemophilus which exhibit immunogenic properties
XX
PS   Claim 3; Figure 2; 100pp; English.
XX
CC   The isolation and purification of the high molecular weight protein
CC   enables the identification of the major protective epitopes of the
CC   protein by conventional epitope mapping.  These epitopes can then be
CC   synthesised using standard techniques and incorporated into fully
CC   synthetic or recombinant vaccines.
XX
SQ   Sequence   1536 AA;
```

alignment_scores:

```
          Quality:  272.50           Length:   1182
            Ratio:  0.452              Gaps:     60
Percent Similarity:  51.015   Percent Identity: 20.897
```

alignment_block:

```
US-09-303-518D-649 x AAR41723   ..

Align seg 1/1  to: AAR41723  from: 1  to: 1536

 127 GGACACACTTATTTCGGCATCAACTACCAATACTATCGCGACTTTGCCGA 176
      |||:::||||||:::|||                        ::::
 349 GlyGluThrTyrLeuGly.........................GlyAs 356

 177 AAATAAAGGCAAGTTTGCAGTCGGGGCGAAAGATATTGAGGTTTACAACA 226
      :::::::|||:::            |||    :::    |||::::::   :::|
 356 pGluArgGlyGlu.........GlyLysAsnGlyIleGlnLeuAlaLysL 370

 227 AAAAAGGGGAGTTGGTCGGCAAATCAATGACAAAAGCCCCGATGATTGAT 276
      ||                         |||:::    |||:::     |||:::
 370 ys.................ThrSerLeuGluLysGlySerThrIleAsn 380

 277 TTTTCTGTGGTGTCGCGTAACGGCGTGGCGGCATTGGTGGGCGATCAATA 326
         |||        :::::::::|||    |||    :::    ||||||
 381 ValSerGlyLysGluLysGlyGlyArgAlaIleValTrpGlyAspIleAl 397

 327 TATTGTG............................AGCGTGGCACATA 346
       ::::::                            ::::::|||    :
 397 aLeuIleAspGlyAsnIleAsnAlaGlnGlySerGlyAspIleAlaLysT 414

 347 ACGGCGGC....................................... 354
      ::||||||
 414 hrGlyGlyPheValGluThrSerGlyHisAspLeuPheIleLysAspAsn 430

 355 ...............................TATAACAACGTTGATTT 371
```

```
                                            :::::||||||::::
 431 AlaIleValAspAlaLysGluTrpLeuLeuAspPheAspAsnValSerIl 447

 372 TGGTGCGGAA......GGAAGAAATCCCGATCAACATCGTTTTACTTAT. 414
     ::::||||||     ||||||:::      ::::::        |||
 447 eAsnAlaGluThrAlaGlyArgSerAsnThrSerGluAspAspGluTyrT 464

 415 ...................AAAATTGTGAAACGGAATAATTATAAAGCA 444
                        :::    |||||||||:::   |||:::
 464 hrGlySerGlyAsnSerAlaSerThrProLysArgAsnLysGluLysThr 480

 445 GGGACTAAAGGCCATCCTTATGGCGGCGATTATCATATGCCGCGTTTGCA 494
          |||           :::   :::|||:::           :::   :
 481 .ThrLeuThrAsnThrThrLeuGluSerIleLeuLysLysGlyThrPheV 497

 495 TAAATTTGTCACAGATGCAGAACCTGTTGAAATGACCAGTTATATGGATG 544
     ::|||::::::       |||:::                |||::::::
 497 alAsnIleThrAlaAsnGlnArg..................IleTyrVal 507

 545 GGCGGAAATATATCGATCAAAATAATTACCCTGACCGTGTTCGTATTGGG 594
     :::::::::|||:::::::::      :::||||||:::::::     ||
 508 AsnSerSerIleAsnLeuSerAsnGlySerLeuThrLeuTrpSerGluGl 524

 595 GCAGGCAGGC..........................AATATTGGCGATC 617
     |::::::|||                           :::   |||||
 524 yArgSerGlyGlyGlyValGluIleAsnAsnAspIleThrThrGlyAspA 541

 618 TGATGAAGATGAGCCCAATAACC.................GCGAAAGTT 649
           :::   :::   :::|||                :::   |||
 541 spThrArgGlyAlaAsnLeuThrIleTyrSerGlyGlyTrpValAspVal 557

 650 CATATCATATTGCAAGTGCGTATTCTTGGCTCGTTGGTGGCAATACCTTT 699
     |||         :::::::::        |||:::::: :::   |||
 558 HisLysAsnIleSerLeuGlyAlaGlnGlyAsnIle..AsnIleThrAla 573

 700 GCACAAAAT...............GGATCA...............GGTGG 719
        |||:::                ||||||               |||
 574 LysGlnAspIleAlaPheGluLysGlySerAsnGlnValIleThrGlyGl 590

 720 TGGCACAGTCAACTTAGGTAGTGAAAAAATTAAACATAGCCCATATGGTT 769
       ||||||:::::   |||:::::|||                  ||||
 590 nGlyThrIleThrSerGlyAsnGlnLys..................GlyP 601

 770 TTTTACCAACAGGAGGCTCATTTGGCGACAGTGGCTCACCAATGTTTATC 819
     ||      ::::::   |||::::::   :::||||||   :::   :::
 601 heArgPheAsnAsnValSerLeuAsnGlyThrGlySerGlyLeuGlnPhe 617

 820 TATGATGCCCAAAAGCAAAAGTGGTTAATTAATGGGGTATTGCAAACGGG 869
                 :::   :::|||:::   |||::::::   ::::::
 618 ThrThrLysArgThrAsnLysTyrAlaIleThrAsnLysPheGluGlyTh 634

 870 CAACCCCTATATAGGAAAAAGCAAT...GGCTTCCAGCTGGTTCGTAAAG 916
                    ||||||   |||   ::::::   |||   ::::::
 634 rLeuAsnIleSerGlyLysValAsnIleSerMetValLeuProLysAsnG 651

 917 ATTGGTTCTATGATGAAATCTTTGCTGGAGATACCCAT......TCAGTA 960
     ::         ||||||:::   |||   |||   |||:::         :::
 651 luSerGlyTyrAspLys...PheLysGlyArgThrTyrTrpAsnLeuThr 666

 961 TTCTACGAACCACGTCAAAATGGGAAATACTCTTTTAACGACGATAATAA 1010
           :::        ::::::|||::::::::::::::::   |||::::::
 667 SerLeuAsnValSerGluSerGlyGluPheAsnLeuThrIleAspSerAr 683

1011 TGGCACA.........GGAAAAATCAATGCCAAACATGAACACAATTCTC 1051
     :|||:::            |||   ::::::      ::::::   |||::::
 683 gGlySerAspSerAlaGlyThrLeuThrGlnProTyrAsnLeuAsnGlyI 700

1052 TGCCT...AATAGATTAAAAACACGAACCGTTCAATTGTTTAATGTTTCT 1098
     ::         |||:::      |||   :::|||:::      |||:::
 700 leSerPheAsnLysAspThrThrPheAsnValGluArg...AsnAlaArg 715

1099 TTATCCGAGACAGCAAGAGAACCTGTTTATCATGCTGCAGGTGGTGTCAA 1148
     ::::::         :::   |||:::               |||:::|||
 716 ValAsnPheAspIleLysAlaProIle...............GlyIleAs 727

1149 CAGTTATCGACCCAGACTGAATAATGGAGAAAATATTTCCTTTATTGACG 1198
     |:::|||                    ::::::::::::::   :::
 727 nLysTyr.....................SerSerLeuAsnTyrAlaSer. 736

1199 AAGGAAAAGGCGAATTGATACTTACCAGCAACATCAATCAAGGTGCTGGA 1248
                             :::::::::||||||:::   ::::::|||
 737 ....................PheAsnGlyAsnIleSerValSerGlyGly 746

1249 GGATTATATTTCCAAGGAGATTTTACGGTCTCGCCTGAAAATAACGAAAC 1298
     |||         :::   ||||||||||:::      ::::::::::::::
 747 Gly.........SerValAspPheThrLeuLeuAlaSerSerSerAsnVa 760

1299 TTGGCAAGGCGCGGGCGTT.....................CATATCAGTG 1327
     :   |||      ||||||                      ::::::|||
 760 l...GlnThrProGlyValValIleAsnSerLysTyrPheAsnValSerT 776

1328 AAGACAGTACCGTTACTTGGAAAGTAAACGGC.................. 1359
        |||::::::   :::|||::::::|||
 776 hrGlySerSerLeuArgPheLysThrSerGlySerThrLysThrGlyPhe 792

1360 ...GTGGCAAACGAC...CGCCTGTCCAAAATCGGCAAAGGCACGCTGCA 1403
        :::   :::|||      |||:::      |||   :::
 793 SerIleGluLysAspLeuThrLeuAsnAlaThrGlyGlyAsnIleThrLe 809

1404 CGTTCAAGCCAAAGGGGAAAAC...........................C 1426
     :::|||::::::|||   :::                               :
 809 uLeuGlnValGluGlyThrAspGlyMetIleGlyLysGlyIleValAlaL 826

1427 AAGGCTCGATCAGCGTGGGCGACGGTACAGTCATTTTGGATCAGCAGGCA 1476
     ::   :::|||:::            |||::::::   :::
 826 ysLysAsnIleThrPheGluGlyGlyAsnIleThrPhe............ 838

1477 GACGATAAAGGCAAAAAACAAGCCTTTAGTGAAATCGGCTTGGTCAGCGG 1526
              |||::::::::::|||   :::||||||         :::||
 839 .........GlySerArgLysAlaValThrGluIle.........GluGl 849

1527 CAGGGGTACGGTGCAACTGAATGCCGATAATCAGTTCAACCCCGACAAAC 1576
     |:::   |||::::::   ||||||:::
 849 yAsnValThrIleAsnAsnAsnAlaAsn...................... 858

1577 TCTATTTCGGCTTTCGCGGCGGACGTTTGGATTTAAACGGGCATTCGCTT 1626
                             :::   |||   |||   ::::::
 859 .........................ValThrLeuIleGlySerAspPhe 866

1627 TCGTTCCACCGTATTCAAAATACCGATGAAGGGGCGATGATTGTCAACCA 1676
     :::   |||:::         |||   :::      :::|||:::|||
 867 AspAsnHisGlnLysProLeuThrIleLysLysAspValIleIleAsnSe 883

1677 CAATCAA......GACAAAGAATCCACCGTTACCATTACAGGCAATAAAG 1720
      ::::::           :::   |||:::|||:::||||||
 883 rGlyAsnLeuThrAlaGlyGlyAsnIleValAsnIleAlaGlyAsnLeuT 900

1721 ATATTGCTACAACCGGCAAT.............................. 1740
        :::   :::::::::|||
 900 hrValGluSerAsnAlaAsnPheLysAlaIleThrAsnPheThrPheAsn 916

1741 ...............AACAACAGCTTGGATAGCAAAAAAGAAATTGCCTA 1775
                   :::|||:::   :::|||:::   :::||||||
 917 ValGlyGlyLeuPheAspAsnLysGlyAsnSerAsnIleSerIleAlaLy 933

1776 CAACGGT...TGGTTTGGCGAGAAAGATACGACCAAAACGAACGGGCGGC 1822
      :::|||      |||   :::   |||::::::|||         :::|
 933 sGlyGlyAlaArgPheLysAspIleAspAsnSerLys.........AsnL 947

1823 TCAACCTTGTTTACCAGCCCGCCGCAGAAGACCGCACCCTGCTGCTTTCC 1872
     ||:::::::::   :::   ::::::         ||||||::::::   :::
 947 euSerIleThrThrAsnSerSerSerThrTyrArgThrIleIleSerGly 963

1873 GGCGGAACAAATTTAAACGGCAACATCACGCAAACAAACGGCAAACTGTT 1922
     :::   ||||||   ||||||:::::::::   ||||||
 964 AsnIleThrAsnLysAsnGlyAspLeuAsnIleThrAsn........... 976
```

```
1923 TTTCAGCGGCAGACCAACACCGCACGCCTACAATCATTTAAACGACCATT 1972
         :::|||      |||                :::::::::
 977 ....GluGlySerAspThrGlu.........MetGlnIleGlyGlyAspV 989

1973 GGTCGCAAAAAGAGGGCATTCCTCGCGGGGAAATCGTGTGGGACAACGAC 2022
       |||||||||||                |||:::::::::   ::::::|||
 989 alSerGlnLysGlu............GlyAsnLeuThrIleSerSerAsp 1001

2023 TGGATCAACCGCACATTTAAAGCGGAAAACTTCCAAATTAAAGGCGGACA 2072
       ||||||   |||            :::::::::   ||||||:::|||
1002 LysIleAsnIleThr.........LysGlnIleThrIleLysAlaGly.. 1014

2073 GGCGGTGGTTTCCCGCAATGTTGCCAAAGTGAAAGGCGATTGGCATTTGA 2122
                         |||      :::::::|||           :
1015 ..................ValAspGlyGluAsnSerAspSerAspAlaT 1025

2123 GCAATCACGCCCAAGCAGTTTTTGGTGTCGCACCGCATCAAAGCCACACA 2172
     ::|||:::|||:::                                |||
1025 hrAsnAsnAlaAsnLeu............................Thr 1031

2173 ATCTGTACACGTTCGGACTGGACGGGTCTGACAAATTGTGTCGAAAAAAC 2222
     |||   |||:::                               |||
1032 IleLysThrLys............................GluLeuLy 1038

2223 CATTACCGACGATAAAGTGATTGCTTCATTGACTAAGACCGACATCAGCG 2272
       :::|||:::|||      |||::::::::::::|||:::::::|||::::
1038 sLeuThrGlnAspLeuAsnIleSerGlyPheAsnLysAlaGluIleThrA 1055

2273 GCAATGTCGATCTTGCCGATCACGCTCATTTAAATCTCACAGGGCTTGCC 2322
     :::::   |||   :::|||
1055 laLys...AspGlySerAsp.........................Leu 1061

2323 ACACTCAACGGCAATCTTAGTGCAAATGGCGATACACGTTATACAGTCAG 2372
     |||::::::::::::   |||||||:::|||   :::           |||::
1062 ThrIleGlyAsnThrAsnSerAlaAspGlyThrAsnAlaLysLysValTh 1078

2373 CCACAACGCCACCCAAAACGGCAACCTTAGCCTCGTGGGCAATGCCCAAG 2422
     :   |||   ::::::::::::::::::|||        |||
1078 rPheAsnGlnValLysAspSerLysIleSerAlaAspGly.......... 1091

2423 CAACATTTAATCAAGCCACATTAAACGGCAACACATCGGCTTCGGGCAAT 2472
              ::::::::||||||:::::::::::::::::|||||||::::
1092 .......HisLysValThrLeuHisSerLysValGluThrSerGlySer 1105

2473 GCTTCATTTAATCTAAGCGACCACGCCGTACAAAACGGCAGTCTGACGCT 2522
        :::   |||   :::|||    :::    :::|||::::::|||||||::
1106 AsnAsn...AsnThrGluAspSerSerAspAsnAsnAlaGlyLeuThrIl 1121

2523 TTCCGGCAACGCTAAGGCAAACGTAAGCCATTCCGCACTCAACGGTAATG 2572
     :      :::||||||   ||||||:::                :::|||:::|||:
1121 e......AspAlaLys...AsnValThr.........ValAsnAsnAsnI 1132

2573 TCTCCCTAGCCGATAAGGCAGTATTCCATTTTGAAAGCAGCCGCTTTACC 2622
     :::::   :::   |||||||||          :::   |||      :::
1132 leThr...SerHisLysAlaVal.........SerIleSerAlaThrSer 1144

2623 GGACAAATCAGCGGCGGCAAGGATACGGCATTACACTTAAAA......GA 2666
     |||:::|||:::                  |||:::::::::           ::
1145 GlyGluIleThrThrLysThrGlyThrThrIleAsnAlaThrThrGlyAs 1161

2667 CAGCGAATGGACGCTGCCGTCAGGCACGGAATTAGGCAATTTAAACCTTG 2716
     :   |||   |||         :::|||:::   ||||||:::::::::     :
1161 nValGluIleThrAlaGlnThrGlySerIleLeuGlyGlyIleGluSerS 1178

2717 ACAACGCCACCATTACACTCAATTCCGCCTATCGCCACGATGCGGCAGGG 2766
     :::::::::::::|||||||:::::::::               |||::::::
1178 erSerGlySerValThrLeuThrAlaThrGluGlyAlaLeuAlaValSer 1194

2767 GCGCAAACCGGCAGTGCGACAGATGCGCCGCGCCGCCGTTCGCGCCGTTC 2816
           :::|||:::::::::   :::      :::   :::      ::
1195 AsnIleSerGlyAsnThrValThrValThrAlaAsnSerGlyAlaLeuTh 1211

2817 GCGCCGTTCCCTATTATCCGTTACACCGCCAACTTCGGTAGAATCCCGTT 2866
     :      :::      ::::::                ||||||   :::
1211 rThrLeuAlaGlySerThrIleLysGlyThrGluSerValThrThrSerS 1228

2867 TCAACACGCTGACGGTAAACGGCAAATTGAACGGTCAGGGAACA...... 2910
     ::::::         :::::::|||   :::::::|||   ||||||
1228 erGlnSerGlyAspIleGlyGlyThrIleSerGly...GlyThrValGlu 1243

2911 TTCCGCTTTATGTCGGAACTCTTCGGCTACCGCAGCGACAAATTGAAGCT 2960
        :::      ::::::|||               ::::::|||:::|||
1244 ValLysAlaThrGluSerLeuThrThrGlnSerAsnSerLysIleLys.. 1259

2961 GGCGGAAAGTTCCGAAGGCACTTACACCTTGGCGGTC............. 2997
       |||   ::::::|||::::::   |||   |||:::
1260 .AlaThrThrGlyGluAlaAsnValThrSerAlaThrGlyThrIleGlyG 1276

2998 ...........AACAATACCGGCAACGAACCTGCAAGCCTCGAACAATTG 3036
                :::||||||   |||      |||:::      :::|||
1276 lyThrIleSerGlyAsnThrValAsnValThrAlaAsnAlaGlyAspLeu 1292

3037 ACGGTA.......................GTGGAAGGAAAAGACAACAA 3062
     ||||||                         :::||||||      :::
1293 ThrValGlyAsnGlyAlaGluIleAsnAlaThrGluGlyAlaAlaThrLe 1309

3063 ACCGCTGTCCGAAAACCTTAATTTCACCCTGCAAAACGAACACGTCGATG 3112
              |||::::::         |||      ::::::|||:::   :
1309 uThrThrSerSerGlyLysLeuThrThrGluAlaSerSerHisIleThrS 1326

3113 CCGGCGCGTGGCGTTACCAACTCATCCGCAAAGACGGCGAGTTCCGCCTG 3162
      :: |||:::|||   :::|||   ||||||:::   :::|||   |||
1326 er.AlaLysGlyGlnValAsnLeuSerAlaGlnAspGlySerValAlaGl 1342

3163 CATAATCCGGTCAAAGAACAAGAGCTTTCCGACAAACTCGGCAAGGCAGA 3212
          |||::::::   |||   ::::::   |||:::::::::
1342 ySerIleAsnAlaAlaAsnValThrLeuAsnThrThrGlyThrLeuThrT 1359

3213 AGCCAAAAAACAGGCGGAAAAAGACAACGCGCAAAGCCTTG 3253
           |||:::   :::   ::::::|||      :::|||
1359 hrValLysGlySerAsnIleAsnAlaThrSerGlyThrLeu 1372
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:AAW30293

```
seq_documentation_block:
ID   AAW30293 standard; Protein; 1536 AA.
XX
AC   AAW30293;
XX
DT   14-APR-1998  (first entry)
XX
DE   Non-typeable Haemophilus high mol.wt. surface protein HMW1.
XX
KW   Non-typeable Haemophilus; high molecular weight surface protein;
KW   HMW1; hmw1A gene; immunogen; vaccine; otitis media.
XX
OS   Haemophilus influenzae strain 12.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 4
FT                   /note= "encoded by CTA"
FT   Misc-difference 98
FT                   /note= "encoded by GAT"
FT   Misc-difference 363
FT                   /note= "encoded by AAG"
XX
PN   WO9736914-A1.
XX
PD   09-OCT-1997.
XX
PF   01-APR-1997;   97WO-US04707.
XX
PR   01-APR-1996;   96US-0617697.
XX
PA   (BARE/) BARENKAMP S J.
```

```
XX
PI   Barenkamp SJ;
XX
DR   WPI; 1997-503038/46.
DR   N-PSDB; AAT90994 and AAT90996.
XX
PT   High molecular weight proteins of non-typeable Haemophilus
PT   influenzae - useful for vaccine production
XX
PS   Claim 7; Page 66-70; 183pp; English.
XX
CC   This protein comprises the high molecular weight surface protein
CC   HMW1 (125 kDa) of non-typeable Haemophilus influenzae strain 12 that
CC   has the immunological ability to protect against disease caused by a
CC   non-typeable Haemophilus strain and is characterised by at least
CC   one surface-exposed B-cell epitope that is recognised by monoclonal
CC   antibody AD6.  The HMW1 amino acid sequence was deduced from the
CC   hmw1 gene sequence (see AAT90994 and AAT90996).  The expressed protein
CC   is truncated, starting at residue 442 of the full-length gene
CC   product.  HMW2 (see AAW30294), HMW3 (see AAW30291) and HMW4 (see
CC   AAW30292) have also been identified.  A conjugate comprising HMW1
CC   linked to an antigen, hapten or polysaccharide, and a synthetic
CC   peptide of 6-150 amino acids corresponding to at least protective
CC   epitope of HMW1 are also claimed.  HMW proteins, conjugates and
CC   peptides can be used in vaccines, as immunogens for preparation of
CC   antibodies and as antigens for detection of these antibodies.
XX
SQ   Sequence   1536 AA;

alignment_scores:
           Quality:  269.50            Length:   1190
             Ratio:  0.452               Gaps:     60
 Percent Similarity:  50.084   Percent Identity: 20.840

alignment_block:
 US-09-303-518D-649 x AAW30293   ..

 Align seg 1/1  to: AAW30293  from: 1  to: 1536

     127 GGACACACTTATTTCGGCATCAACTACCAATACTATCGCGACTTTGCCGA 176
         |||:::||||||:::|||                          :::::
     349 GlyGluThrTyrLeuGly..........................GlyAs 356

     177 AAATAAAGGCAAGTTTGCAGTCGGGGCGAAAGATATTGAGGTTTACAACA 226
         :::::::|||:::             |||   :::   |||::::::   :::|
     356 pGluArgGlyGlu.........GlyLysAsnGlyIleGlnLeuAlaLysL 370

     227 AAAAAGGGGAGTTGGTCGGCAAATCAATGACAAAAGCCCCGATGATTGAT 276
         ||                    |||:::   |||:::      |||:::
     370 ys..................ThrSerLeuGluLysGlySerThrIleAsn 380

     277 TTTTCTGTGGTGTCGCGTAACGGCGTGGCGGCATTGGTGGGCGATCAATA 326
           |||       :::::::::|||   |||   :::   ||||||
     381 ValSerGlyLysGluLysGlyGlyArgAlaIleValTrpGlyAspIleAl 397

     327 TATTGTG............................AGCGTGGCACATA 346
          ::::::                            :::::::|||   :
     397 aLeuIleAspGlyAsnIleAsnAlaGlnGlySerGlyAspIleAlaLysT 414

     347 ACGGCGGCTAT.................................... 357
         ::||||||:::
     414 hrGlyGlyPheValGluThrSerGlyHisAspLeuPheIleLysAspAsn 430

     358 ..................................AACAACGTTGATTT 371
                                            :::|||||||::::
     431 AlaIleValAspAlaLysGluTrpLeuLeuAspProAspAsnValSerIl 447

     372 TGGTGCGGAA......GGAAGAAATCCCGATCAACATCGTTTTACTTAT. 414
         ::::||||||        ||||||:::     ::::::        |||
     447 eAsnAlaGluThrAlaGlyArgSerAsnThrSerGluAspAspGluTyrT 464

     415 ...................AAAATTGTGAAACGGAATAATTATAAAGCA 444
                              :::         ||||||||||:::   |||:::
     464 hrGlySerGlyAsnSerAlaSerThrProLysArgAsnLysGluLysThr 480

     445 GGGACTAAAGGCCATCCTTATGGCGGCGATTATCATATGCCGCGTTTGCA 494
           |||               :::   :::|||:::          :::   :
     481 .ThrLeuThrAsnThrThrLeuGluSerIleLeuLysLysGlyThrPheV 497

     495 TAAATTTGTCACAGATGCAGAACCTGTTGAAATGACCAGTTATATGGATG 544
         ::|||:::::       |||:::                   |||::::::
     497 alAsnIleThrAlaAsnGlnArg.................IleTyrVal 507

     545 GGCGGAAATATATCGATCAAAATAATTACCCTGACCGTGTTCGTATTGGG 594
         :::::::::|||:::::::::      :::||||||::::::      ||
     508 AsnSerSerIleAsnLeuSerAsnGlySerLeuThrLeuTrpSerGluGl 524

     595 GCAGGCAGGC.........................AATATTGGCGATC 617
         |::::::|||                         :::   ||||||
     524 yArgSerGlyGlyGlyValGluIleAsnAsnAspIleThrThrGlyAspA 541

     618 TGATGAAGATGAGCCCAATAACC.................GCGAAAGTT 649
            :::   :::   :::|||                  :::   |||
     541 spThrArgGlyAlaAsnLeuThrIleTyrSerGlyGlyTrpValAspVal 557

     650 CATATCATATTGCAAGTGCGTATTCTTGGCTCGTTGGTGGCAATACCTTT 699
         |||     :::::::::        |||::::::  :::   |||
     558 HisLysAsnIleSerLeuGlyAlaGlnGlyAsnIle..AsnIleThrAla 573

     700 GCACAAAAT..............GGATCA..............GGTGG 719
            |||:::              ||||||              |||
     574 LysGlnAspIleAlaPheGluLysGlySerAsnGlnValIleThrGlyGl 590

     720 TGGCACAGTCAACTTAGGTAGTGAAAAAATTAAACATAGCCCATATGGTT 769
          ||||||:::::   |||::::::|||                  ||||
     590 nGlyThrIleThrSerGlyAsnGlnLys.................GlyP 601

     770 TTTTACCAACAGGAGGCTCATTTGGCGACAGTGGCTCACCAATGTTTATC 819
         ||       ::::::   |||::::::   :::||||||   :::   :::
     601 heArgPheAsnAsnValSerLeuAsnGlyThrGlySerGlyLeuGlnPhe 617

     820 TATGATGCCCAAAAGCAAAAGTGGTTAATTAATGGGGTATTGCAAACGGG 869
                  :::   :::|||:::   |||::::::   ::::::
     618 ThrThrLysArgThrAsnLysTyrAlaIleThrAsnLysPheGluGlyTh 634

     870 CAACCCCTATATAGGAAAAAGCAAT...GGCTTCCAGCTGGTTCGTAAAG 916
                        ||||||   |||   ::::::   |||   :::::::
     634 rLeuAsnIleSerGlyLysValAsnIleSerMetValLeuProLysAsnG 651

     917 ATTGGTTCTATGATGAAATCTTTGCTGGAGATACCCAT......TCAGTA 960
          ::       |||||||:::   |||   |||   |||:::           :::
     651 luSerGlyTyrAspLys...PheLysGlyArgThrTyrTrpAsnLeuThr 666

     961 TTCTACGAACCACGTCAAAATGGGAAATACTCTTTTAACGACGATAATAA 1010
              :::       ::::::|||:::::::::::::::   |||::::::
     667 SerLeuAsnValSerGluSerGlyGluPheAsnLeuThrIleAspSerAr 683

    1011 TGGCACA.........GGAAAAATCAATGCCAAACATGAACACAATTCTC 1051
          :|||:::         |||   ::::::   ::::::   |||:::
     683 gGlySerAspSerAlaGlyThrLeuThrGlnProTyrAsnLeuAsnGlyI 700

    1052 TGCCT...AATAGATTAAAAACACGAACCGTTCAATTGTTTAATGTTTCT 1098
          ::        |||:::      |||   :::|||:::      |||:::
     700 leSerPheAsnLysAspThrThrPheAsnValGluArg...AsnAlaArg 715

    1099 TTATCCGAGACAGCAAGAGAACCTGTTTATCATGCTGCAGGTGGTGTCAA 1148
          ::::::       :::   |||:::                |||:::||
     716 ValAsnPheAspIleLysAlaProIle..............GlyIleAs 727

    1149 CAGTTATCGACCCAGACTGAATAATGGAGAAAATATTTCCTTTATTGACG 1198
          |:::|||                    ::::::::::::::   :::
     727 nLysTyr....................SerSerLeuAsnTyrAlaSer. 736

    1199 AAGGAAAAGGCGAATTGATACTTACCAGCAACATCAATCAAGGTGCTGGA 1248
                                  :::::::::||||||:::   ::::::|||
```

```
 737 ....................PheAsnGlyAsnIleSerValSerGlyGly 746

1249 GGATTATATTTCCAAGGAGATTTTACGGTCTCGCCTGAAAATAACGAAAC 1298
     |||          :::   |||||||||:::      ::::::::::::::
 747 Gly.........SerValAspPheThrLeuLeuAlaSerSerSerAsnVa 760

1299 TTGGCAAGGCGCGGGCGTT.....................CATATCAGTG 1327
     :   |||      ||||||                      :::::::|||
 760 l...GlnThrProGlyValValIleAsnSerLysTyrPheAsnValSerT 776

1328 AAGACAGTACCGTTACTTGGAAAGTAAACGGC.................. 1359
         |||:::::   :::|||::::::|||
 776 hrGlySerSerLeuArgPheLysThrSerGlySerThrLysThrGlyPhe 792

1360 ...GTGGCAAACGAC...CGCCTGTCCAAAATCGGCAAAGGCACGCTGCA 1403
        :::   :::|||     |||:::      |||   :::
 793 SerIleGluLysAspLeuThrLeuAsnAlaThrGlyGlyAsnIleThrLe 809

1404 CGTTCAAGCCAAAGGGGAAAAC...........................C 1426
      :::|||:::::|||    :::                          :
 809 uLeuGlnValGluGlyThrAspGlyMetIleGlyLysGlyIleValAlaL 826

1427 AAGGCTCGATCAGCGTGGGCGACGGTACAGTCATTTTGGATCAGCAGGCA 1476
     ::   :::|||:::          |||:::::::   :::
 826 ysLysAsnIleThrPheGluGlyGlyAsnIleThrPhe............ 838

1477 GACGATAAAGGCAAAAAACAAGCCTTTAGTGAAATCGGCTTGGTCAGCGG 1526
                |||:::::::|||   :::||||||          :::||
 839 .........GlySerArgLysAlaValThrGluIle.........GluGl 849

1527 CAGGGGTACGGTGCAACTGAATGCCGATAATCAGTTCAACCCCGACAAAC 1576
     |:::    |||:::::   |||||||:::
 849 yAsnValThrIleAsnAsnAsnAlaAsn......................: 858

1577 TCTATTTCGGCTTTCGCGGCGGACGTTTGGATTTAAACGGGCATTCGCTT 1626
                                :::   |||   |||   ::::::
 859 .........................ValThrLeuIleGlySerAspPhe 866

1627 TCGTTCCACCGTATTCAAAATACCGATGAAGGGGCGATGATTGTCAACCA 1676
     :::   |||:::           |||    :::      :::|||:::|||
 867 AspAsnHisGlnLysProLeuThrIleLysLysAspValIleIleAsnSe 883

1677 CAATCAA......GACAAAGAATCCACCGTTACCATTACAGGCAATAAAG 1720
      ::::::                  :::   |||:::|||:::||||||
 883 rGlyAsnLeuThrAlaGlyGlyAsnIleValAsnIleAlaGlyAsnLeuT 900

1721 ATATTGCTACAACCGGCAAT.............................. 1740
       :::   :::::::|||
 900 hrValGluSerAsnAlaAsnPheLysAlaIleThrAsnPheThrPheAsn 916

1741 ...............AACAACAGCTTGGATAGCAAAAAAGAAATTGCCTA 1775
                    :::|||:::   :::|||:::   :::||||||
 917 ValGlyGlyLeuPheAspAsnLysGlyAsnSerAsnIleSerIleAlaLy 933

1776 CAACGGT...TGGTTTGGCGAGAAAGATACGACCAAAACGAACGGGCGGC 1822
      :::|||       |||   :::   |||:::::::|||         :::|
 933 sGlyGlyAlaArgPheLysAspIleAspAsnSerLys.........AsnL 947

1823 TCAACCTTGTTTACCAGCCCGCCGCAGAAGACCGCACCCTGCTGCTTTCC 1872
     ||:::::::::   :::   ::::::         ||||||::::::   :::
 947 euSerIleThrThrAsnSerSerSerThrTyrArgThrIleIleSerGly 963

1873 GGCGGAACAAATTTAAACGGCAACATCACGCAAACAAACGGCAAACTGTT 1922
     :::   ||||||   ||||||:::::::::   ||||||
 964 AsnIleThrAsnLysAsnGlyAspLeuAsnIleThrAsn........... 976

1923 TTTCAGCGGCAGACCAACACCGCACGCCTACAATCATTTAAACGACCATT 1972
         :::|||       |||            :::::::::
 977 ....GluGlySerAspThrGlu.........MetGlnIleGlyGlyAspV 989

1973 GGTCGCAAAAAGAGGGCATTCCTCGCGGGGAAATCGTGTGGGACAACGAC 2022
       |||||||||||             |||:::::::::   :::::::|||
 989 alSerGlnLysGlu............GlyAsnLeuThrIleSerSerAsp 1001

2023 TGGATCAACCGCACATTTAAAGCGGAAAACTTCCAAATTAAAGGCGGACA 2072
       ||||||   |||          :::::::::   ||||||:::|||
1002 LysIleAsnIleThr.........LysGlnIleThrIleLysAlaGly.. 1014

2073 GGCGGTGGTTTCCCGCAATGTTGCCAAAGTGAAAGGCGATTGGCATTTGA 2122
                       |||      :::::::|||            :
1015 ...................ValAspGlyGluAsnSerAspSerAspAlaT 1025

2123 GCAATCACGCCCAAGCAGTTTTTGGTGTCGCACCGCATCAAAGCCACACA 2172
     ::|||:::|||:::                                |||
1025 hrAsnAsnAlaAsnLeu.............................Thr 1031

2173 ATCTGTACACGTTCGGACTGGACGGGTCTGACAATTGTGTCGAAAAAAC 2222
     |||   |||:::                                  |||
1032 IleLysThrLys...........................GluLeuLy 1038

2223 CATTACCGACGATAAAGTGATTGCTTCATTGACTAAGACCGACATCAGCG 2272
      :::|||:::|||      |||:::::::::::|||::::::|||:::
1038 sLeuThrGlnAspLeuAsnIleSerGlyPheAsnLysAlaGluIleThrA 1055

2273 GCAATGTCGATCTTGCCGATCACGCTCATTTAAATCTCACAGGGCTTGCC 2322
     :::::   |||   :::|||
1055 laLys...AspGlySerAsp.........................Leu 1061

2323 ACACTCAACGGCAATCTTAGTGCAAATGGCGATACACGTTATACAGTCAG 2372
     |||::::::::::::   ||||||:::|||    :::       |||::
1062 ThrIleGlyAsnThrAsnSerAlaAspGlyThrAsnAlaLysLysValTh 1078

2373 CCACAACGCCACCCAAAACGGCAACCTTAGCCTCGTGGGCAATGCCCAAG 2422
     :   |||   ::::::::::::::::::|||      |||:::      :
1078 rPheAsnGlnValLysAspSerLysIleSerAlaAspGlyHisLysValT 1095

2423 CAACATTTAATCAAGCCACATTAAACGGCAACACATCGGCTTCGGGCAAT 2472
     ::      :::::::::   :::|||:::::::::   :::   :::
1095 hrLeuHisSerLysValGluThrSerGlySerAsnAsnAsnThrGluHis 1111

2473 GCTTCATTTAATCTAAGCGACCACGCCGTACAAAACGGCAGTCTGACGCT 2522
     :::|||   |||   :::      :::::::::      ::::::|||::
1112 SerSerAspAsnAsnAlaGlyLeuThrIleAspAlaLysAsnValThrVa 1128

2523 TTCCGGCAACGCTAAGGCAAAC......GTAAGCCATTCCGCACTCAACG 2566
     :::::::|||      ::::::       ||||||   ||||||   :::|
1128 lAsnAsnAsnIleThrSerHisLysAlaValSerIleSerAlaThrSerG 1145

2567 GTAATGTC...............TCCCTAGCCGATAAGGCAGTATTCCAT 2601
     ||::::::                :::      :::   ::::::   :::
1145 lyGluIleThrThrLysThrGlyThrThrIleAsnAlaThrThrGlyAsn 1161

2602 TTTGAAAGCAGCCGCTTTACCGGACAAATCAGCGGCGGCAAGGATACG.. 2649
        |||   :::      ||||||:::|||   ||||||   ::::::
1162 ValGluIleThrAlaGlnThrGlySerIleLeuGlyGlyIleGluSerSe 1178

2650 .......GCATTACACTTAAAAGACAGCGAATGGACGCTGCCGTCAGGCA 2692
        ::::::   |||      :::|||   :::|||
1178 rSerGlySerValThrLeuThrAlaThrGluGlyAlaLeu.......... 1191

2693 CGGAATTAGGCAATTTAAACCTTGACAACGCCACCATTACACTCAATTCC 2742
        ::::::|||::::::   :::::::::|||:::|||   ||||||
1192 ..AlaValSerAsnIleSerGlyAsnThrValThrValThrAlaAsnSer 1207

2743 GCCTATCGCCACGATGCGGCAGGGGCGCAAACCGGCAGTGCGACAGATGC 2792
     :::                  ||||||:::   |||   ::::::|||:::
1208 GlyAlaLeuThrThrLeuAlaGlySer...ThrIleLysGlyThrGlu.. 1222

2793 GCCGCGCCGCCGTTCGCGCCGTTCGCGCCGTTCCCTATTATCCGTTACAC 2842
                                             |||||||||
1223 ......................................SerValThr. 1225

2843 CGCCAACTTCGGTAGAATCCCGTTTCAACACGCTGACGGTAAACGGCAAA 2892
          ||||||   :::|||                  :::::::|||
1226 .....ThrSerSerGlnSerGly...........AspIleGlyGlyThr 1236
```

```
2893 TTGAACGGTCAGGGAACA......TTCCGCTTTATGTCGGAACTCTTCGG 2936
     :::::|||   ||||||          :::       :::::|||
1237 IleSerGly...GlyThrValGluValLysAlaThrGluSerLeuThrTh 1252

2937 CTACCGCAGCGACAAATTGAAGCTGGCGGAAAGTTCCGAAGGCACTTACA 2986
           ::::::|||:::|||     |||   :::::|||::::::    |
1252 rGlnSerAsnSerLysIleLys...AlaThrThrGlyGluAlaAsnValT 1268

2987 CCTTGGCGGTC.......................AACAATACCGGCAAC 3012
      ||    |||:::                        :::||||||   |||
1268 hrSerAlaThrGlyThrIleGlyGlyThrIleSerGlyAsnThrValAsn 1284

3013 GAACCTGCAAGCCTCGAACAATTGACGGTA.................... 3042
             |||:::        :::|||||||||
1285 ValThrAlaAsnAlaGlyAspLeuThrValGlyAsnGlyAlaGluIleAs 1301

3043 ....GTGGAAGGAAAAGACAACAAACCGCTGTCCGAAAACCTTAATTTCA 3088
         :::||||||        :::            |||::::::          |
1301 nAlaThrGluGlyAlaAlaThrLeuThrThrSerSerGlyLysLeuThrT 1318

3089 CCCTGCAAAACGAACACGTCGATGCCGGCGCGTGGCGTTACCAACTCATC 3138
      ||      ::::::|||:::   :::  |||:::|||    :::|||   ||
1318 hrGluAlaSerSerHisIleThrSer.AlaLysGlyGlnValAsnLeuSe 1334

3139 CGCAAAGACGGCGAGTTCCGCCTGCATAATCCGGTCAAAGAACAAGAGCT 3188
      ||||:::    :::|||    |||        |||::::::    |||    ::::
1334 rAlaGlnAspGlySerValAlaGlySerIleAsnAlaAlaAsnValThrL 1351

3189 TTCCGACAAACTCGGCAAGGCAGAAGCCAAAAAACAGGCGGAAAAAGACA 3238
      ::    |||:::::::::               |||:::    :::    ::::::
1351 euAsnThrThrGlyThrLeuThrThrValLysGlySerAsnIleAsnAla 1367

3239 ACGCGCAAAGCCTTG 3253
      |||       :::|||
1368 ThrSerGlyThrLeu 1372
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01837

seq_documentation_block:

```
ID   AAB01837 standard; Protein; 1073 AA.
XX
AC   AAB01837;
XX
DT   11-SEP-2000  (first entry)
XX
DE   Haemophilus influenzae strain LCDC2 mature HMW2A protein, SEQ ID NO:49.
XX
KW   Mature HMW protein; hmw gene; hmwA1; hmwA2; high molecular weight;
KW   non-typeable Haemophilus influenzae; NTHi; non-encapsulated;
KW   recombinant production; Escherichia coli; antibacterial; vaccine;
KW   human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;
KW   detection; diagnosis.
XX
OS   Haemophilus influenzae strain LCDC2.
XX
PN   WO200020609-A2.
XX
PD   13-APR-2000.
XX
PF   07-OCT-1999;   99WO-CA00938.
XX
PR   07-OCT-1998;   98US-0167568.
PR   08-DEC-1998;   98US-0206942.
XX
PA   (CONN-) CONNAUGHT LAB LTD.
XX
PI   Loosmore SM,  Yang Y,  Klein MH;
XX
DR   WPI; 2000-303789/26.
DR   N-PSDB; AAA52186.
XX
PT   Nucleic acid molecule for producing recombinant high molecular weight
PT   proteins of Haemophilus which are used as a vaccine to provide
PT   protection against Haemophilus induced diseases in humans -
XX
PS   Claim 8; Fig 23A-P; 307pp; English.
XX
CC   The invention relates to the recombinant production of Haemophilus
CC   influenzae high molecular weight (HMW) proteins in Escherichia coli. The
CC   expression construct used to effect recombinant expression comprises a
CC   promoter functional in E. coli (e.g., the T7 promoter) operably linked
CC   to a modified hmwABC operon from a non-typeable (non-encapsulated) H.
CC   influenzae (NTHi). Most HMW-expressing NTHi strains contain two hmw gene
CC   clusters termed hmw1ABC and hmw2ABC. Each hmwABC operon comprises hmwA,
CC   hmwB and hmwC genes. The hmwA genes encode the structural HMWA proteins
CC   and the hmwB and hmwC genes encode accessory proteins which are
CC   responsible for post-translational processing and secretion of the HMWA
CC   proteins. The modified hmwABC operon used in the expression construct of
CC   the invention contains an A gene modified such that it encodes only the
CC   mature HMWA. The invention also discloses hmwA genes (AAA52175-A52198)
CC   and HMWA proteins (AAB01824-B01849) from the non-typeable H. influenzae
CC   strains Joyc, K1, K21, LCDC2, PMH1, 15 and 12. The nucleic acids and
CC   vectors are used for the production of recombinant H. influenzae HMW
CC   proteins which can be used as vaccines to mediate a humoral or
CC   cell-mediated immune response to provide protection against diseases in
CC   humans caused by H. influenzae (e.g., otitis media, epiglottitis,
CC   pneumonia and tracheobronchitis). The HMW proteins are also useful as
CC   antigens in immunoassays for detecting antibodies against Haemophilus,
CC   HMW proteins and/or HMW peptides. The nucleotide sequences encoding the
CC   HMW proteins can be used to isolate and clone hmw genes from other
CC   non-typeable strains of Haemophilus via hybridisation reactions. The
CC   present sequence represents a mature HMWA protein from a non-typeable
CC   strain of H. influenzae.
XX
SQ   Sequence   1073 AA;
```

alignment_scores:

```
          Quality:  268.00           Length:   1062
            Ratio:  0.485              Gaps:     53
Percent Similarity:  52.072   Percent Identity: 20.716
```

alignment_block:

```
US-09-303-518D-649 x AAB01837   ..

Align seg 1/1  to: AAB01837  from: 1  to: 1073

 787 TCATTTGGCGACAGTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCA 836
      ::::::||||||||||||       :::    :::|||       :::|||::
  72 AsnIleGlyAspSerGlyHis...LeuThrLeuTyrLysLysArgLysAs 87

 837 AAAGTGGTTAATTAATGGGGTATTGCAAACGGGCAACCCCTATATAGGAA 886
      ::::          :::     :::|||   ::::::          :::
  87 nArg.........SerAspGlyIleGlnIleAsnLysAspIleThrSerT 101

 887 AAAGCAATGGCTTCCAGCTGGTTCGTAAAGATTGG......TTCTATGAT 930
       ::::::::::::   :::           ||||||      ::::::
 101 hrGlyGlySerLeuThrIleAsnSerAspAspTrpValAspIleHisGly 117

 931 GAAATCTTTGCTGGAGAT.....................ACCCATTCAGT 959
      :::|||      |||:::                     :::   |||||
 118 AsnIleThrLeuGlyGluGlyPheLeuAsnIleThrSerSerAspSerVa 134

 960 ATTCTACGAA........................................ 969
      |   :::|||
 134 lAlaPheGluGlyGlyAsnGlyAsnLysGlyArgSerSerAlaSerAlaG 151

 970 ............................CCACGTCAAAATGGGAAATAC 990
                                     :::|||      :::
 151 lnIleIleAlaGlnGlyThrIleThrLeuThrGlyGluAsnLysThrPhe 167

 991 TCTTTTAACGACGATAAT...AATGGCACAGGAAAA...ATCAATGCCAA 1034
       :::|||:::    :::    |||||||||||:::     ::::::
 168 ArgLeuAsnAsnValSerLeuAsnGlyThrGlyAsnGlyLeuSerIleIl 184

1035 ACATGAACACAATTCTCTGCCTAATAGATTAAAAACACGAACCGTTCAAT 1084
```

```
                      :::::|||    :::||||||          :::
 184 eSerThrAlaSerAsnLeuSerHisArgLeuAspGlyGluIleAsnValS 201

1085 TGTTTAATGTTTCTTTATCCGAGACAGCAAGAGAACCTGTT...TATCAT 1131
          |||||::::::::::::|||:::::::::    :::    |||
 201 erGlyAsnValThrIleAsnGlnThrThrGlnGlnAsnIleGluTyrTrp 217

1132 GCTGCAGGT...............GGTGTCAACAGTTATCGACCCAGA.. 1164
        |||:::                 :::|||:::||||:::::   |||
 218 LysAlaSerSerAspSerTyrTrpAsnValThrSerPheAsnLeuArgGl 234

1165 ...........................CTGAATAATGGAGAAAAT.... 1182
                                 :::|||::::::::::|||
 234 uAspSerLysPheThrPheIleLysTyrValAsnSerAlaArgAsnGlyA 251

1183 ....................ATTTCCTTTATTGACGAAGGAAAAGGCGAA 1212
                           :::   |||    :::::||||||
 251 spValArgGlyArgSerPheAlaGlyValIlePheAsnAlaLysGly... 266

1213 TTGATACTTACCAGCAACATCAATCAAGGTGCTGGAGGATTATATTTCCA 1262
           ||||||:::::::::|||            |||:::
 267 ......LeuThrThrSerPheAsnValLysLysGlySerThrVal..... 279

1263 AGGAGATTTTACGGTCTCGCCTGAAAATAACGAAACTTGGCAAGGCGCGG 1312
         ||||||    ::::::|||:::::::::    :::   |||
 280 ....AspPheLysLeuLysProAsnSerGlyTyrAsnSerGln...LysA 294

1313 GCGTTCATATCAGTGAAGACAGTACCGTTACTTGGAAAGTAAACGGCGTG 1362
       :::    |||:::    :::|||:::::::::    :::    :::|||
 294 rgIleProIleGlnPheGlnSerAsnIleSerValSerGlyGlyGlyArg 310

1363 GCAAAC...GACCGCCTGTCCAAAATCGGCAAAGGCACGCTGCACGTTCA 1409
     :::|||     :::     |||:::::::::      |||     ::::::::::::
 311 ValAsnIleAsnThrLeuAlaAsnLeuThrGlyGlyGlyValGluIleAr 327

1410 AGCCAAAGGGGAAAACCAAGGCTCGATCAGCGTGGGCGACGGT...ACAG 1456
     ::::                   :::||||||:::|||:::||||||    |||:
 327 gSer...............SerSerIleAsnValSerAspGlySerThrL 339

1457 TCATTTTGGATCAGCAGGCAGACGATAAAGGCAAAAAACAAGCCTTTAGT 1506
     ::   :::          ||||||   |||::::::      :::      :::
 339 euSerMetThrAlaGlnAlaArgAspArgAsnAlaPheGluIleThrLys 355

1507 GAAATCGGCTTGGTCAGCGGCAGGGGTACGGTGCAACTG........... 1545
     ::::::   :::   ::::::::::::::::::::::::
 356 AspLeuValIleAsnAlaSerAsnSerAsnLeuSerIleIleGlnGlnAs 372

1546 ...........................AATGCCGATAATCAGTTCAACC 1567
                                 ||||||    |||
 372 nAspGlyPheAspAsnAsnGlnLysAlaAsnAlaIleAsn.......... 385

1568 CCGACAAACTCTATTTCGGCTTTCGCGGCGGACGTTTGGATTTAAACGGG 1617
        :::|||               :::::||||||:::::   |||:::|||
 386 ..SerLysTyrAsnValThrIleGlnGlyGlyAsnValThrLeuGlyGly 401

1618 CATTCGCTTTCGTTCCACCGTATTCAAAATACCGATGAAGGGGCGATGAT 1667

 401 .................................................. 401

1668 TGTCAACCACAATCAAGACAAAGAATCCACCGTTACCATTACAGGCAATA 1717
                           ::::::|||:::    ||||||||||||:::
 402 ..................GlnAsnSerSerSerThrIleThrGlySerV 412

1718 AAGATATTGCTACAACCGGCAAT...............AACAACAGCTTG 1752
       :::|||::::::::::::|||                  :::|||:::
 412 alAsnIleGlyAlaAsnAlaAsnValThrLeuGlnAlaHisAsnGlyAsn 428

1753 GATAGCAAAAAAGAAATTGCCTACAACGGTTGGTTTGGCGAGAAAGATAC 1802
     |||    :::|||::::::::::::::                      ::
 429 AspArgAsnLysLysLeuThrPheGly.....................As 438

1803 GACCAAAACGAACGGGCGGCTCAACCTTGTTTACCAGCCCGCCGCAGAAG 1852
     ::::::::::::|||:::|||:::||||||          |||       :
```

```
 438 nValSerValGluGlyGluLeuArgLeuValGlyAlaSerAlaAsnIleA 455

1853 ACCGCACCCTGCTGCTTTCCGGCGGAACAAATTTAAACGGCAACATCACG 1902
     ::::::::|||    ::::::::|||:::::::::::::          :::
 455 snAsnAsnLeuSerValLysSerGlyAlaLysPheLys.........Ala 468

1903 CAAACAAACGGCAAACTGTTTTTCAGCGGCAGACCAACACCGCACGCCTA 1952
     ::::||||||    :::|||    ::::::|||       |||    ::::::
 469 GluThrAsnAspAsnLeuAsnIleThrGlyThrPheThrAsnAsnGlyTh 485

1953 CAATCATTTAAACGACCATTGGTCGCAAAAAGAGGGCATTCCTCGC...G 1999
      :::   ::::::                      |||:::|||      :::   |
 485 rSerIleIleAsp............ValLysLysGlyAlaAlaLysLeuG 498

2000 GGGAAATCGTGTGGGACAACGACTGGATCAACCGCACATTTAAAGCGGAA 2049
     ||:::|||:::    |||:::::::    :::|||    |||    :::|||:::
 498 lyAsnIleThrAsnAspGlyAsn...LeuAsnIleThrThrAsnAlaLys 513

2050 AACTTCCAAATTAAAGGCGGACAGGCGGTGGTTTCCCGCAATGTTGCCAA 2099
     |||             |||    ::::::|||:::::    |||:::::::::
 514 Asn............GlyGlnLysSerValIleAsnGlyAsnIleThrAs 526

2100 AGTGAAAGGCGATTGGCATTTGAGCAATCACGCCCAA............G 2137
     :    ||||||      :::::::::|||:::::::::
 526 nAsnLysGlyAlaLeuAsnIleThrAsnAsnGlyAsnAspThrGluIleG 543

2138 CAGTTTTTGGTGTCGCACCGCATCAAAGCCACACAATCTGTACACGTTCG 2187
       :::    |||             :::::::::   ::::::          |||
 543 lnIleGlyGlyAsnIleSerGlnLysGluGlyAsnLeuThrIleSerSer 559

2188 GACTGGACGGGTCTGACAAATTGTGTCGAA...AAAACCATTACCGAC.. 2232
     |||       ::::::|||:::    :::|||    |||:::    ||||||
 560 AspLysIleAsnIleThrLysArgIleGluIleLysAlaGlyThrAspGl 576

2233 .........GATAAAGTGATTGCTTCATTGACTAAGACCGACATCAGCG 2272
                  |||:::    :::||||||    ::::::       |||:::
 576 nGlyAsnSerAspSerGlyValAlaSerAsnAlaAsnLeuThrIleLys. 592

2273 GCAATGTCGATCTTGCCGATCACGCTCATTTAAATCTCACAGGG...... 2316
       :::    :::|||                :::||||||::::::|||
 593 ..ThrLysGluLeuLysLeuThrGluAsnLeuAsnIleSerGlyPheAsp 608

2317 ...........CTTGCCACACTCAACGGCAATCTT.........AGTGC 2345
                :::|||       |||:::||||||               :::
 609 LysAlaGluIleValAlaLysGluAsnAsnAsnLeuIleIleGlyAsnAs 625

2346 AAATGGCGAT......ACACGTTATACAGTCAGCCACAACGCCACCCAAA 2389
      |||||||||        :::          ||||||:::    |||    ::::::
 625 nAsnGlyAspAsnAlaAsnAlaLysThrValThrPheAsnAsnValLysA 642

2390 ACGGCAACCTTAGCCTCGTGGGCAATGCCCAAGCAACATTTAATCAA... 2436
     :::::::::|||       |||:::    ::::::|||:::|||:::
 642 spSerLysIleSerAlaAsnGlyHis...AsnValThrLeuAsnSerLys 657

2437 ...GCCACATTAAACGGCAACACATCGGCTTCGGGCAATGCTTCATTTAA 2483
           |||    :::||||||::::::::::::|||||||::::::    ||
 658 ValGluThrSerAspGlyAsnSerAsnThrGluGlyAsnSerAspAsnAs 674

2484 T.........................................CTAA 2488
     |                                           ::::
 674 nAlaGlyLeuThrIleAspAlaLysAsnValThrValAsnAsnAspIleT 691

2489 GCGACCACGCCGTACAAAACGGCAGTCTGACGCTTTCCGGCAACGCTAAG 2538
     ::::::|||    :::    |||    :::    :::         ::::::|||
 691 hrSerHisLysThrValAsnIleThrAlaSerGluArgIleAspThrLys 707

2539 GCAAACGTAAGCCATTCCGCACTCAACGGTAATGTCTCCCTA........ 2580
     |||:::::::::    :::|||    :::||||||||||:::|||
 708 AlaAspThrThrIleAsnAlaThrThrGlyAsnValLysLeuThrAlaVa 724

2581 .........................GCCGATAAGGCAGTATTCCATT 2602
                                :::::::::::        ::::
 724 lThrSerAspIleGlnGlyGlyIleLysSerAsnSerGlyAspValAsnI 741
```

```
2603 TTGAAAGCAGCCGCTTTACCGGACAAATCAGCGGC.........GGCAAG 2643
     ::   :::|||         ||||||:::|||:::|||          ::::|||
 741 leThrThrSer......ThrGlySerIleAsnGlyLysIleGluSerLys 755

2644 GATACGGCATTACACTTAAAAGACAGCGAATGGACGCTGCCGTCAGGCAC 2693
     :::   ::::::    |||       :::|||   ||||||
 756 SerGlySerValThrLeuThrAlaThrGluLysThrLeu........... 768

2694 GGAATTAGGCAATTTAAACCTTGACAACGCCACCATTACACTCAATTCCG 2743
        :::|||||||:::::    :::::::::|||:::|||    |||    :
 769 .ThrValGlyAsnValSerGlyAsnThrValThrValThrAlaAsnArgG 785

2744 CCTATCGCCACGATGCGGCAGGGGCGCAAACCGGCAGTGCGACAGATGCG 2793
     ::                   |||||||:::    |||    ::::::|||::::::
 785 lyAlaLeuThrThrLeuAlaGlySer...ThrIleAsnGlyThrAsnGly 800

2794 CCGCGCCGCCGTTCGCGCCGTTCGCGCCGTTCCCTATTATCCGTTACACC 2843
                   |||:::    ::::::    :::          :::
 801 ValThrThrSerSerGlnSerGlyGluIleGlyGlyGluValThrGlyLy 817

2844 GCCAACTTCGGTAGAATCCCGTTTCAACACGCTGACGGTAAACGGC.... 2889
        :::||||||     :::        ::::::|||||||||:::|||
 817 sThrValSerValThrAlaThrAlaGlySerLeuThrValLysGlyGlyA 834

2890 ..AAATTGAACGGT...CAGGGAACATTCCGCTTTATGTCGGAACTCTTC 2934
       |||:::|||:::    :::||||||        :::   ::::::
 834 laLysIleAsnAlaThrGluGlyThrAlaThrLeuThrAlaSer...... 848

2935 GGCTACCGCAGCGACAAATTGAAGCTGGCGGAAAGTTCCGAAGGCACT.. 2982
                   |||   ||||||           ||||||:::   |||
 849 .........SerGlyLysLeuThrThrGluAlaSerSerAsnIleThrSe 862

2982 .................................................. 2982

 862 rAlaLysGlyGlnValAspLeuSerAlaGlnAspGlySerIleAlaGlyG 879

2983 ..........TACACCTTGGCGGTCAACAATACCGGCAACGAACCTGCA 3021
                  ::::::::::::|||:::|||||||:::
 879 lnIleSerAlaAlaAsnValThrLeuAsnThrThrGlyThr......... 892

3022 AGCCTCGAACAATTGACGGTAGTGGAAGGAAAAGACAACAAACCGCTGTC 3071
                      ||||||:::|||||||||  ::::::    |||
 893 ...........LeuThrThrValGluGly.SerSerIleAsnAlaAsnG 905

3072 CGAAAACCTTAATTTCACCCTGC.....................AAAACG 3100
               ||||||:::     :::                    |||:::
 905 luGlyThrLeuValIleAsnAlaAsnAspAlaLysLeuAspGlyLysAla 921

3101 AACACGTCGATGCCGGCGCGTGGCGTTACCAACTCATCC........... 3139
     :::    :::              ::::::|||||||:::|||
 922 SerGlyAsnArgThrGluValAsnAlaThrAsnAlaSerGlySerGlySe 938

3140 .......GCAAAGACGGCGAGTTCCGCCTGCATAATCCGGTCAAAGAACA 3182
               |||||||||:::|||||||:::   |||       :::   |||
 938 rValThrAlaLysThrSerSerSerValAsnIleThrGlyAspLeuAsnT 955

3183 AGAGCTTTCCGACAAACTCGGCAAGGCAGAAGCCAAAAAACAGGCGGAAA 3232
                        |||          ::::::   :::|||       :::
 955 hrIleAsnGlyLeuAsnIleIleSerGluAsnGlyArgAsnThrValArg 971

3233 AAGACAACGCGCAAAGCCTTGACGCGCTGATTGCGGCCGGGCGCGAT... 3279
              ::::::       ::::::   ::::::||||||:::
 972 LeuArgGlyLysGluIleGluValLysTyrIleGlnProGlyValAlaSe 988

3280 ...........................GCCGTCGAAAAGACAGAAAGCG 3301
                               ::::::||||||:::::::::
 988 rValGluGluValIleGluAlaLysArgValLeuGluLysValLysAspL 1005

3302 TTGCCGAACCGGCCCGGCAGGCAGGCGGGGAAAATGTCGGCATTATGCAG 3351
     ::::::::        |||::::::     ::::::    :::|||:::
1005 euSerAspGluGluArgGluThrLeuAlaLys...LeuGlyValSerAla 1020

3352 GCGGAGGAAGAGAAAAAACGGGTGCAGGCGGATAAAGACACCGCCTTGGC 3401
     :::              :::    ::::::    ::::::       ::::::
1021 Val............ArgPheIleGluProAsnAsnThrIleThrValAs 1033

3402 GAAACAGCGCGAAGCGGAAACCCGGCCGGCTACC 3435
          |||:::|||       |||||||||::::::
1033 nThrGlnAsnGluPheThrThrArgProSerSer 1044
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01836

```
seq_documentation_block:
ID   AAB01836 standard; Protein; 1079 AA.
XX
AC   AAB01836;
XX
DT   11-SEP-2000  (first entry)
XX
DE   Haemophilus influenzae strain LCDC2 HMW2A protein, SEQ ID NO:47.
XX
KW   HMW protein; hmw gene; hmwA1; hmwA2; high molecular weight;
KW   non-typeable Haemophilus influenzae; NTHi; non-encapsulated;
KW   recombinant production; Escherichia coli; antibacterial; vaccine;
KW   human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;
KW   detection; diagnosis.
XX
OS   Haemophilus influenzae strain LCDC2.
XX
PN   WO200020609-A2.
XX
PD   13-APR-2000.
XX
PF   07-OCT-1999;   99WO-CA00938.
XX
PR   07-OCT-1998;   98US-0167568.
PR   08-DEC-1998;   98US-0206942.
XX
PA   (CONN-) CONNAUGHT LAB LTD.
XX
PI   Loosmore SM,  Yang Y,  Klein MH;
XX
DR   WPI; 2000-303789/26.
DR   N-PSDB; AAA52185.
XX
PT   Nucleic acid molecule for producing recombinant high molecular weight
PT   proteins of Haemophilus which are used as a vaccine to provide
PT   protection against Haemophilus induced diseases in humans -
XX
PS   Claim 12; Fig 23A-P; 307pp; English.
XX
CC   The invention relates to the recombinant production of Haemophilus
CC   influenzae high molecular weight (HMW) proteins in Escherichia coli. The
CC   expression construct used to effect recombinant expression comprises a
CC   promoter functional in E. coli (e.g., the T7 promoter) operably linked
CC   to a modified hmwABC operon from a non-typeable (non-encapsulated) H.
CC   influenzae (NTHi). Most HMW-expressing NTHi strains contain two hmw gene
CC   clusters termed hmw1ABC and hmw2ABC. Each hmwABC operon comprises hmwA,
CC   hmwB and hmwC genes. The hmwA genes encode the structural HMWA proteins
CC   and the hmwB and hmwC genes encode accessory proteins which are
CC   responsible for post-translational processing and secretion of the HMWA
CC   proteins. The modified hmwABC operon used in the expression construct of
CC   the invention contains an A gene modified such that it encodes only the
CC   mature HMWA. The invention also discloses hmwA genes (AAA52175-A52198)
CC   and HMWA proteins (AAB01824-B01849) from the non-typeable H. influenzae
CC   strains Joyc, K1, K21, LCDC2, PMH1, 15 and 12. The nucleic acids and
CC   vectors are used for the production of recombinant H. influenzae HMW
CC   proteins which can be used as vaccines to mediate a humoral or
CC   cell-mediated immune response to provide protection against diseases in
CC   humans caused by H. influenzae (e.g., otitis media, epiglottitis,
CC   pneumonia and tracheobronchitis). The HMW proteins are also useful as
CC   antigens in immunoassays for detecting antibodies against Haemophilus,
CC   HMW proteins and/or HMW peptides. The nucleotide sequences encoding the
CC   HMW proteins can be used to isolate and clone hmw genes from other
CC   non-typeable strains of Haemophilus via hybridisation reactions. The
```

```
CC   present sequence represents an HMWA protein from a non-typeable strain of
CC   H. influenzae.
XX
SQ   Sequence   1079 AA;

alignment_scores:
           Quality:  268.00          Length:   1062
             Ratio:  0.485             Gaps:     53
 Percent Similarity: 52.072   Percent Identity: 20.716

alignment_block:
 US-09-303-518D-649 x AAB01836   ..

 Align seg 1/1  to: AAB01836  from: 1  to: 1079

  787 TCATTTGGCGACAGTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCA 836
      :::::||||||||||||       :::   :::|||      :::|||::
   78 AsnIleGlyAspSerGlyHis...LeuThrLeuTyrLysLysArgLysAs 93

  837 AAAGTGGTTAATTAATGGGGTATTGCAAACGGGCAACCCCTATATAGGAA 886
       ::::        :::     :::|||   ::::::        :::
   93 nArg.........SerAspGlyIleGlnIleAsnLysAspIleThrSerT 107

  887 AAAGCAATGGCTTCCAGCTGGTTCGTAAAGATTGG......TTCTATGAT 930
        :::::::::::   :::        ||||||      ::::::
  107 hrGlyGlySerLeuThrIleAsnSerAspAspTrpValAspIleHisGly 123

  931 GAAATCTTTGCTGGAGAT....................ACCCATTCAGT 959
      :::|||      |||:::                     :::   |||||
  124 AsnIleThrLeuGlyGluGlyPheLeuAsnIleThrSerSerAspSerVa 140

  960 ATTCTACGAA....................................... 969
      |   :::|||
  140 lAlaPheGluGlyGlyAsnGlyAsnLysGlyArgSerSerAlaSerAlaG 157

  970 ..........................CCACGTCAAAATGGGAAATAC 990
                                     :::|||       :::
  157 lnIleIleAlaGlnGlyThrIleThrLeuThrGlyGluAsnLysThrPhe 173

  991 TCTTTTAACGACGATAAT...AATGGCACAGGAAAA...ATCAATGCCAA 1034
         :::|||:::   :::   ||||||||||||:::   ::::::
  174 ArgLeuAsnAsnValSerLeuAsnGlyThrGlyAsnGlyLeuSerIleIl 190

 1035 ACATGAACACAATTCTCTGCCTAATAGATTAAAAACACGAACCGTTCAAT 1084
                ::::::|||   :::||||||      :::
  190 eSerThrAlaSerAsnLeuSerHisArgLeuAspGlyGluIleAsnValS 207

 1085 TGTTTAATGTTTCTTTATCCGAGACAGCAAGAGAACCTGTT...TATCAT 1131
         ||||||:::::::::::|||:::::::::   :::   |||
  207 erGlyAsnValThrIleAsnGlnThrThrGlnGlnAsnIleGluTyrTrp 223

 1132 GCTGCAGGT..............GGTGTCAACAGTTATCGACCCAGA.. 1164
         |||:::                 :::|||:::|||::::::   |||
  224 LysAlaSerSerAspSerTyrTrpAsnValThrSerPheAsnLeuArgGl 240

 1165 ..........................CTGAATAATGGAGAAAAT.... 1182
                                 :::|||::::::::::|||
  240 uAspSerLysPheThrPheIleLysTyrValAsnSerAlaArgAsnGlyA 257

 1183 ....................ATTTCCTTTATTGACGAAGGAAAAGGCGAA 1212
                          :::    |||    ::::::||||||
  257 spValArgGlyArgSerPheAlaGlyValIlePheAsnAlaLysGly... 272

 1213 TTGATACTTACCAGCAACATCAATCAAGGTGCTGGAGGATTATATTTCCA 1262
            ||||||:::::::::|||            |||:::
  273 ......LeuThrThrSerPheAsnValLysLysGlySerThrVal..... 285

 1263 AGGAGATTTTACGGTCTCGCCTGAAAATAACGAAACTTGGCAAGGCGCGG 1312
           ||||||    ::::::|||:::::::::    :::    |||
  286 ....AspPheLysLeuLysProAsnSerGlyTyrAsnSerGln...LysA 300

 1313 GCGTTCATATCAGTGAAGACAGTACCGTTACTTGGAAAGTAAACGGCGTG 1362
       :::    |||:::    :::|||:::::::::    :::    :::|||
  300 rgIleProIleGlnPheGlnSerAsnIleSerValSerGlyGlyGlyArg 316

 1363 GCAAAC...GACCGCCTGTCCAAAATCGGCAAAGGCACGCTGCACGTTCA 1409
      :::|||    :::    |||:::::::::       |||    ::::::::::
  317 ValAsnIleAsnThrLeuAlaAsnLeuThrGlyGlyGlyValGluIleAr 333

 1410 AGCCAAAGGGGAAAACCAAGGCTCGATCAGCGTGGGCGACGGT...ACAG 1456
      ::::                  :::||||||:::|||:::||||||    |||:
  333 gSer..............SerSerIleAsnValSerAspGlySerThrL 345

 1457 TCATTTTGGATCAGCAGGCAGACGATAAAGGCAAAAAACAAGCCTTTAGT 1506
      ::    :::         ||||||   |||::::::      :::       :::
  345 euSerMetThrAlaGlnAlaArgAspArgAsnAlaPheGluIleThrLys 361

 1507 GAAATCGGCTTGGTCAGCGGCAGGGGTACGGTGCAACTG........... 1545
      ::::::    :::   ::::::::::::::::::::::
  362 AspLeuValIleAsnAlaSerAsnSerAsnLeuSerIleIleGlnGlnAs 378

 1546 ..........................AATGCCGATAATCAGTTCAACC 1567
                                ||||||    |||
  378 nAspGlyPheAspAsnAsnGlnLysAlaAsnAlaIleAsn.......... 391

 1568 CCGACAAACTCTATTTCGGCTTTCGCGGCGGACGTTTGGATTTAAACGGG 1617
         :::|||          ::::::||||||::::::    |||:::|||
  392 ..SerLysTyrAsnValThrIleGlnGlyGlyAsnValThrLeuGlyGly 407

 1618 CATTCGCTTTCGTTCCACCGTATTCAAAATACCGATGAAGGGGCGATGAT 1667

  407 ................................................ 407

 1668 TGTCAACCACAATCAAGACAAAGAATCCACCGTTACCATTACAGGCAATA 1717
                        :::::::|||:::   ||||||||||||:::
  408 ..................GlnAsnSerSerSerThrIleThrGlySerV 418

 1718 AAGATATTGCTACAACCGGCAAT...............AACAACAGCTTG 1752
        :::|||::::::::::::|||                  :::|||:::
  418 alAsnIleGlyAlaAsnAlaAsnValThrLeuGlnAlaHisAsnGlyAsn 434

 1753 GATAGCAAAAAAGAAATTGCCTACAACGGTTGGTTTGGCGAGAAAGATAC 1802
      |||    :::|||::::::::::::::::                  ::
  435 AspArgAsnLysLysLeuThrPheGly....................As 444

 1803 GACCAAAACGAACGGGCGGCTCAACCTTGTTTACCAGCCCGCCGCAGAAG 1852
      :::::::::::::|||:::|||:::||||||          |||       :
  444 nValSerValGluGlyGluLeuArgLeuValGlyAlaSerAlaAsnIleA 461

 1853 ACCGCACCCTGCTGCTTTCCGGCGGAACAAATTTAAACGGCAACATCACG 1902
      :::::::::|||    :::::::::|||::::::::::::         :::
  461 snAsnAsnLeuSerValLysSerGlyAlaLysPheLys.........Ala 474

 1903 CAAACAAACGGCAAACTGTTTTTCAGCGGCAGACCAACACCGCACGCCTA 1952
      :::||||||    :::|||    ::::::|||       |||    ::::::
  475 GluThrAsnAspAsnLeuAsnIleThrGlyThrPheThrAsnAsnGlyTh 491

 1953 CAATCATTTAAACGACCATTGGTCGCAAAAAGAGGGCATTCCTCGC...G 1999
       :::    ::::::               |||:::|||      :::    |
  491 rSerIleIleAsp............ValLysLysGlyAlaAlaLysLeuG 504

 2000 GGGAAATCGTGTGGGACAACGACTGGATCAACCGCACATTTAAAGCGGAA 2049
      |||:::|||:::   |||:::::::   :::|||    |||    :::|||:::
  504 lyAsnIleThrAsnAspGlyAsn...LeuAsnIleThrThrAsnAlaLys 519

 2050 AACTTCCAAATTAAAGGCGGACAGGCGGTGGTTTCCCCGCAATGTTGCCAA 2099
      |||           |||    ::::::|||::::::    |||:::::::::
  520 Asn............GlyGlnLysSerValIleAsnGlyAsnIleThrAs 532

 2100 AGTGAAAGGCGATTGGCATTTGAGCAATCACGCCCAA............G 2137
      :     ||||||         :::::::::|||:::::::::
  532 nAsnLysGlyAlaLeuAsnIleThrAsnAsnGlyAsnAspThrGluIleG 549

 2138 CAGTTTTTGGTGTCGCACCGCATCAAAGCCACACAATCTGTACACGTTCG 2187
        :::    |||          :::::::::    ::::::          |||
```

```
 549 lnIleGlyGlyAsnIleSerGlnLysGluGlyAsnLeuThrIleSerSer 565

2188 GACTGGACGGGTCTGACAAATTGTGTCGAA...AAAACCATTACCGAC.. 2232
     |||      ::::::|||:::    :::|||   |||:::   ||||||
 566 AspLysIleAsnIleThrLysArgIleGluIleLysAlaGlyThrAspGl 582

2233 ..........GATAAAGTGATTGCTTCATTGACTAAGACCGACATCAGCG 2272
                   |||:::   :::||||||   ::::::      |||:::
 582 nGlyAsnSerAspSerGlyValAlaSerAsnAlaAsnLeuThrIleLys. 598

2273 GCAATGTCGATCTTGCCGATCACGCTCATTTAAATCTCACAGGG...... 2316
       :::   :::|||              :::||||||::::::|||
 599 ..ThrLysGluLeuLysLeuThrGluAsnLeuAsnIleSerGlyPheAsp 614

2317 ...........CTTGCCACACTCAACGGCAATCTT.........AGTGC 2345
                :::|||     |||:::||||||          :::
 615 LysAlaGluIleValAlaLysGluAsnAsnAsnLeuIleIleGlyAsnAs 631

2346 AAATGGCGAT......ACACGTTATACAGTCAGCCACAACGCCACCCAAA 2389
      |||||||||        :::      ||||||:::   |||    :::::::
 631 nAsnGlyAspAsnAlaAsnAlaLysThrValThrPheAsnAsnValLysA 648

2390 ACGGCAACCTTAGCCTCGTGGGCAATGCCCAAGCAACATTTAATCAA... 2436
     :::::::::::|||      |||:::   ::::::|||:::|||:::
 648 spSerLysIleSerAlaAsnGlyHis...AsnValThrLeuAsnSerLys 663

2437 ...GCCACATTAAACGGCAACACATCGGCTTCGGGCAATGCTTCATTTAA 2483
          |||    :::||||||:::::::::::::::||||||:::::    ||
 664 ValGluThrSerAspGlyAsnSerAsnThrGluGlyAsnSerAspAsnAs 680

2484 T.........................................CTAA 2488
     |                                          ::::
 680 nAlaGlyLeuThrIleAspAlaLysAsnValThrValAsnAsnAspIleT 697

2489 GCGACCACGCCGTACAAAACGGCAGTCTGACGCTTTCCGGCAACGCTAAG 2538
     ::::::|||   :::    |||   :::    :::          :::::::|||
 697 hrSerHisLysThrValAsnIleThrAlaSerGluArgIleAspThrLys 713

2539 GCAAACGTAAGCCATTCCGCACTCAACGGTAATGTCTCCCTA........ 2580
     |||:::::::::    :::|||    :::||||||||||:::|||
 714 AlaAspThrThrIleAsnAlaThrThrGlyAsnValLysLeuThrAlaVa 730

2581 ..........................GCCGATAAGGCAGTATTCCATT 2602
                               :::::::::::       ::::
 730 lThrSerAspIleGlnGlyGlyIleLysSerAsnSerGlyAspValAsnI 747

2603 TTGAAAGCAGCCGCTTTACCGGACAAATCAGCGGC.........GGCAAG 2643
     ::    :::|||         ||||||:::|||:::|||          :::|||
 747 leThrThrSer......ThrGlySerIleAsnGlyLysIleGluSerLys 761

2644 GATACGGCATTACACTTAAAAGACAGCGAATGGACGCTGCCGTCAGGCAC 2693
     :::   ::::::   |||         :::|||   ||||||
 762 SerGlySerValThrLeuThrAlaThrGluLysThrLeu........... 774

2694 GGAATTAGGCAATTTAAACCTTGACAACGCCACCATTACACTCAATTCCG 2743
       :::||||||:::::    :::::::::|||:::|||    |||    :
 775 .ThrValGlyAsnValSerGlyAsnThrValThrValThrAlaAsnArgG 791

2744 CCTATCGCCACGATGCGGCAGGGGCGCAAACCGGCAGTGCGACAGATGCG 2793
     ::               ||||||:::    |||    ::::::||||::::::
 791 lyAlaLeuThrThrLeuAlaGlySer...ThrIleAsnGlyThrAsnGly 806

2794 CCGCGCCGCCGTTCGCGCCGTTCGCGCCGTTCCCTATTATCCGTTACACC 2843
                  |||:::   ::::::   :::         :::
 807 ValThrThrSerSerGlnSerGlyGluIleGlyGlyGluValThrGlyLy 823

2844 GCCAACTTCGGTAGAATCCCGTTTCAACACGCTGACGGTAAACGGC.... 2889
        :::||||||   :::      ::::::||||||||||||:::|||
 823 sThrValSerValThrAlaThrAlaGlySerLeuThrValLysGlyGlyA 840

2890 ..AAATTGAACGGT...CAGGGAACATTCCGCTTTATGTCGGAACTCTTC 2934
        |||:::|||:::   :::||||||          :::    ::::::
 840 laLysIleAsnAlaThrGluGlyThrAlaThrLeuThrAlaSer...... 854

2935 GGCTACCGCAGCGACAAATTGAAGCTGGCGGAAAGTTCCGAAGGCACT.. 2982
               |||   ||||||            ||||||:::   |||
 855 .........SerGlyLysLeuThrThrGluAlaSerSerAsnIleThrSe 868

2982 .................................................. 2982

 868 rAlaLysGlyGlnValAspLeuSerAlaGlnAspGlySerIleAlaGlyG 885

2983 ...........TACACCTTGGCGGTCAACAATACCGGCAACGAACCTGCA 3021
                :::::::::::|||:::||||||:::
 885 lnIleSerAlaAlaAsnValThrLeuAsnThrThrGlyThr......... 898

3022 AGCCTCGAACAATTGACGGTAGTGGAAGGAAAAGACAACAAACCGCTGTC 3071
                 ||||||:::|||||||||   ::::::     |||
 899 ............LeuThrThrValGluGly.SerSerIleAsnAlaAsnG 911

3072 CGAAAACCTTAATTTCACCCTGC.....................AAAACG 3100
          ||||||:::      :::                        |||:::
 911 luGlyThrLeuValIleAsnAlaAsnAspAlaLysLeuAspGlyLysAla 927

3101 AACACGTCGATGCCGGCGCGTGGCGTTACCAACTCATCC........... 3139
     :::   :::                ::::::|||||||:::|||
 928 SerGlyAsnArgThrGluValAsnAlaThrAsnAlaSerGlySerGlySe 944

3140 .......GCAAAGACGGCGAGTTCCGCCTGCATAATCCGGTCAAAGAACA 3182
              |||||||||:::||||||:::    |||       :::   |||
 944 rValThrAlaLysThrSerSerSerValAsnIleThrGlyAspLeuAsnT 961

3183 AGAGCTTTCCGACAAACTCGGCAAGGCAGAAGCCAAAAAACAGGCGGAAA 3232
                   |||      ::::::   :::|||       :::
 961 hrIleAsnGlyLeuAsnIleIleSerGluAsnGlyArgAsnThrValArg 977

3233 AAGACAACGCGCAAAGCCTTGACGCGCTGATTGCGGCCGGGCGCGAT... 3279
              ::::::      ::::::   ::::::||||||:::
 978 LeuArgGlyLysGluIleGluValLysTyrIleGlnProGlyValAlaSe 994

3280 ...........................GCCGTCGAAAAGACAGAAAGCG 3301
                                ::::::||||||::::::::::
 994 rValGluGluValIleGluAlaLysArgValLeuGluLysValLysAspL 1011

3302 TTGCCGAACCGGCCCGGCAGGCAGGCGGGGAAAATGTCGGCATTATGCAG 3351
     ::::::::        |||::::::   ::::::    :::|||:::
1011 euSerAspGluGluArgGluThrLeuAlaLys...LeuGlyValSerAla 1026

3352 GCGGAGGAAGAGAAAAAACGGGTGCAGGCGGATAAAGACACCGCCTTGGC 3401
     :::          :::   ::::::   ::::::       ::::::
1027 Val...........ArgPheIleGluProAsnAsnThrIleThrValAs 1039

3402 GAAACAGCGCGAAGCGGAAACCCGGCCGGCTACC 3435
          |||:::|||       |||||||||::::::
1039 nThrGlnAsnGluPheThrThrArgProSerSer 1050
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:AAR41725

```
seq_documentation_block:
ID   AAR41725 standard; Protein; 1536 AA.
XX
AC   AAR41725;
XX
DT   26-APR-1994  (first entry)
XX
DE   High molecular weight protein 1 (HMW1).
XX
KW   HMW; high molecular weight protein; virus; vaccine; influenza;
KW   epitope; immunity; haemophilus influenzae; gene cluster.
XX
OS   Haemophilus influenzae.
XX
FH   Key              Location/Qualifiers
FT   Misc-difference  668..677
FT                    /note= "Possibly incorrect sequence. Alternative
FT                           sequence for this region is LNVSESGEFN.
```

```
FT                         (See comments)."
XX
PN   WO9319090-A.
XX
PD   30-SEP-1993.
XX
PF   16-MAR-1993;   93WO-US02166.
XX
PR   16-MAR-1992;   92GB-0005704.
XX
PA   (BARE/) BARENKAMP S J.
PA   (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
XX
PI   Barenkamp SJ;
XX
DR   WPI; 1993-320683/40.
DR   N-PSDB; AAQ49508.
XX
PT   High molecular weight surface proteins - of non-typeable
PT   haemophilus which exhibit immunogenic properties
XX
PS   Claim 3; Figure 2/10; 100pp; English.
XX
CC   The isolation and purification of the high molecular weight protein
CC   enables the identification of the major protective epitopes of the
CC   protein by conventional epitope mapping.  These epitopes can then be
CC   synthesised using standard techniques and incorporated into fully
CC   synthetic or recombinant vaccines.  This sequence is claimed to be
CC   the same as that given in AAR41723 (High molecular weight protein 1)
CC   although it does differ slightly. (Repeated regions which are
CC   possibly incorrect and occur in the corresponding nucleotide coding
CC   sequence contribute to these differences).
XX
SQ   Sequence   1536 AA;


alignment_scores:
          Quality:  267.50          Length:   1173
            Ratio:  0.444             Gaps:     58
 Percent Similarity:  51.407   Percent Identity: 21.057

alignment_block:
 US-09-303-518D-649 x AAR41725   ..

 Align seg 1/1  to: AAR41725  from: 1  to: 1536

     127 GGACACACTTATTTCGGCATCAACTACCAATACTATCGCGACTTTGCCGA 176
         |||::::|||||||:::|||                         :::::
     349 GlyGluThrTyrLeuGly.........................GlyAs 356

     177 AAATAAAGGCAAGTTTGCAGTCGGGGCGAAAGATATTGAGGTTTACAACA 226
         :::::::|||:::            |||   :::   |||::::::   :::|
     356 pGluArgGlyGlu.........GlyLysAsnGlyIleGlnLeuAlaLysL 370

     227 AAAAAGGGGAGTTGGTCGGCAAATCAATGACAAAAGCCCCGATGATTGAT 276
          ||                  |||:::   |||:::      |||:::
     370 ys.................ThrSerLeuGluLysGlySerThrIleAsn 380

     277 TTTTCTGTGGTGTCGCGTAACGGCGTGGCGGCATTGGTGGGCGATCAATA 326
           |||       :::::::::|||   |||    :::    ||||||
     381 ValSerGlyLysGluLysGlyGlyArgAlaIleValTrpGlyAspIleAl 397

     327 TATTGTG............................AGCGTGGCACATA 346
          ::::::                           :::::|||    :
     397 aLeuIleAspGlyAsnIleAsnAlaGlnGlySerGlyAspIleAlaLysT 414

     347 ACGGCGGCTAT.................................... 357
         ::|||||||:::
     414 hrGlyGlyPheValGluThrSerGlyHisAspLeuPheIleLysAspAsn 430

     358 ..................................AACAACGTTGATTT 371
                                          :::||||||:::::
     431 AlaIleValAspAlaLysGluTrpLeuLeuAspProAspAsnValSerIl 447

     372 TGGTGCGGAA......GGAAGAAATCCCGATCAACATCGTTTTACTTAT. 414
          ::::||||||      ||||||:::      ::::::         |||
     447 eAsnAlaGluThrAlaGlyArgSerAsnThrSerGluAspAspGluTyrT 464

     415 ...................AAAATTGTGAAACGGAATAATTATAAAGCA 444
                        :::      |||||||||:::    |||:::
     464 hrGlySerGlyAsnSerAlaSerThrProLysArgAsnLysGluLysThr 480

     445 GGGACTAAAGGCCATCCTTATGGCGGCGATTATCATATGCCGCGTTTGCA 494
             |||            :::    :::|||:::          :::   :
     481 .ThrLeuThrAsnThrThrLeuGluSerIleLeuLysLysGlyThrPheV 497

     495 TAAATTTGTCACAGATGCAGAACCTGTTGAAATGACCAGTTATATGGATG 544
         ::|||::::::      |||:::                  |||::::::
     497 alAsnIleThrAlaAsnGlnArg..................IleTyrVal 507

     545 GGCGGAAATATATCGATCAAAATAATTACCCTGACCGTGTTCGTATTGGG 594
         :::::::::|||:::::::::        :::||||||::::::      ||
     508 AsnSerSerIleAsnLeuSerAsnGlySerLeuThrLeuTrpSerGluGl 524

     595 GCAGGCAGGC..........................AATATTGGCGATC 617
         |::::::|||                          :::   ||||||
     524 yArgSerGlyGlyGlyValGluIleAsnAsnAspIleThrThrGlyAspA 541

     618 TGATGAAGATGAGCCCAATAACC.................GCGAAAGTT 649
             :::   :::   :::|||                  :::    |||
     541 spThrArgGlyAlaAsnLeuThrIleTyrSerGlyGlyTrpValAspVal 557

     650 CATATCATATTGCAAGTGCGTATTCTTGGCTCGTTGGTGGCAATACCTTT 699
         |||        :::::::::         |||::::::   :::    |||
     558 HisLysAsnIleSerLeuGlyAlaGlnGlyAsnIle..AsnIleThrAla 573

     700 GCACAAAAT...............GGATCA...............GGTGG 719
           ||||:::              ||||||                |||
     574 LysGlnAspIleAlaPheGluLysGlySerAsnGlnValIleThrGlyGl 590

     720 TGGCACAGTCAACTTAGGTAGTGAAAAAATTAAACATAGCCCATATGGTT 769
           ||||||::::::   |||::::::|||                ||||
     590 nGlyThrIleThrSerGlyAsnGlnLys..................GlyP 601

     770 TTTTACCAACAGGAGGCTCATTTGGCGACAGTGGCTCACCAATGTTTATC 819
         ||        ::::::   |||::::::   :::||||||   :::    :::
     601 heArgPheAsnAsnValSerLeuAsnGlyThrGlySerGlyLeuGlnPhe 617

     820 TATGATGCCCAAAAGCAAAAGTGGTTAATTAATGGGGTATTGCAAACGGG 869
                :::    :::|||:::   |||::::::    ::::::
     618 ThrThrLysArgThrAsnLysTyrAlaIleThrAsnLysPheGluGlyTh 634

     870 CAACCCCTATATAGGAAAAAGCAAT...GGCTTCCAGCTGGTTCGTAAAG 916
                        ||||||    |||    ::::::    |||    ::::::
     634 rLeuAsnIleSerGlyLysValAsnIleSerMetValLeuProLysAsnG 651

     917 ATTGGTTCTATGATGAAATCTTTGCTGGAGATACCCATTCAGTATTCTAC 966
         ::        ||||||:::   |||   |||   |||:::          :::
     651 luSerGlyTyrAspLys...PheLysGlyArgThrTyrTrpAsnLeuThr 666

     967 GAACCACGTCAA......AATGGGAAATACTCTTTTAACGACGATAATAA 1010
         :::      :::        |||:::|||   :::::::::   |||::::::
     667 SerLysValAspMetIleAsnSerLysAspAlaLeuThrIleAspSerAr 683

    1011 TGGCACA.........GGAAAAATCAATGCCAAACATGAACACAATTCTC 1051
         :|||:::          |||    ::::::      ::::::   |||::::
     683 gGlySerAspSerAlaGlyThrLeuThrGlnProTyrAsnLeuAsnGlyI 700

    1052 TGCCT...AATAGATTAAAAACACGAACCGTTCAATTGTTTAATGTTTCT 1098
         ::         |||:::      |||    :::|||:::       |||:::
     700 leSerPheAsnLysAspThrThrPheAsnValGluArg...AsnAlaArg 715

    1099 TTATCCGAGACAGCAAGAGAACCTGTTTATCATGCTGCAGGTGGTGTCAA 1148
         ::::::          :::   |||:::                |||:::||
     716 ValAsnPheAspIleLysAlaProIle..............GlyIleAs 727
```

```
1149 CAGTTATCGACCCAGACTGAATAATGGAGAAAATATTTCCTTTATTGACG 1198
     |:::|||                         ::::::::::::::  :::
 727 nLysTyr....................SerSerLeuAsnTyrAlaSer. 736

1199 AAGGAAAAGGCGAATTGATACTTACCAGCAACATCAATCAAGGTGCTGGA 1248
                                 ::::::::|||||||:::  :::::|||
 737 ....................PheAsnGlyAsnIleSerValSerGlyGly 746

1249 GGATTATATTTCCAAGGAGATTTTACGGTCTCGCCTGAAAATAACGAAAC 1298
     |||         :::   ||||||||:::       ::::::::::::::
 747 Gly.........SerValAspPheThrLeuLeuAlaSerSerSerAsnVa 760

1299 TTGGCAAGGCGCGGGCGTT.....................CATATCAGTG 1327
     :   |||       ||||||                    ::::::|||
 760 l...GlnThrProGlyValValIleAsnSerLysTyrPheAsnValSerT 776

1328 AAGACAGTACCGTTACTTGGAAAGTAAACGGC.................. 1359
         |||::::::   :::|||:::::::|||
 776 hrGlySerSerLeuArgPheLysThrSerGlySerThrLysThrGlyPhe 792

1360 ...GTGGCAAACGAC...CGCCTGTCCAAAATCGGCAAAGGCACGCTGCA 1403
        :::   :::|||       |||:::      |||    :::
 793 SerIleGluLysAspLeuThrLeuAsnAlaThrGlyGlyAsnIleThrLe 809

1404 CGTTCAAGCCAAAGGGGAAAACCAAGGCTCGATCAGCGTGGGCGACGGTA 1453
      :::|||:::::::|||     :::                :::|||   |||
 809 uLeuGlnValGluGlyThrAspGly.........MetIleGlyLysGlyI 823

1454 CAGTCATTTTGGATCAGCAGGCAGACGATAAAGGCAAAAAACAAGCCTTT 1503
        |||         :::   :::   :::   ||||||   :::
 823 leValAlaLysLysAsnIleThrPheGluGlyGlyLysMetArgPheGly 839

1504 AGTGAAATCGGCTTGGTCAGCGGCAGGGGTACGGTGCAACTGAATGCCGA 1553
     |||:::   ::::::::::::   :::|||:::|||   :::|||
 840 SerArgLysAlaValThrGluIleGluGlyAsnValThrIleAsn..... 854

1554 TAATCAGTTCAACCCCGACAAACTCTATTTCGGCTTTCGCGGCGGACGTT 1603
                                            ::::::::::::
 855 ....................................AsnAsnAlaAsnV 859

1604 TGGATTTAAACGGGCATTCGCTTTCGTTCCACCGTATTCAAAATACCGAT 1653
     ::   |||   |||   :::::::::   |||:::          |||
 859 alThrLeuIleGlySerAspPheAspAsnHisGlnLysProLeuThrIle 875

1654 GAAGGGGCGATGATTGTCAACCACAATCAA......GACAAAGAATCCAC 1697
     :::       :::|||:::|||     ::::::                :::
 876 LysLysAspValIleIleAsnSerGlyAsnLeuThrAlaGlyGlyAsnIl 892

1698 CGTTACCATTACAGGCAATAAAGATATTGCTACAACCGGCAAT....... 1740
       |||:::|||:::||||||       :::   :::::::|||
 892 eValAsnIleAlaGlyAsnLeuThrValGluSerAsnAlaAsnPheLysA 909

1741 ......................................AACAACAGCTTG 1752
                                            :::|||:::
 909 laIleThrAsnPheThrPheAsnValGlyGlyLeuPheAspAsnLysGly 925

1753 GATAGCAAAAAAGAAATTGCCTACAACGGT...TGGTTTGGCGAGAAAGA 1799
     :::|||:::   :::||||||   :::|||      |||   :::   ||
 926 AsnSerAsnIleSerIleAlaLysGlyGlyAlaArgPheLysAspIleAs 942

1800 TACGACCAAAACGAACGGGCGGCTCAACCTTGTTTACCAGCCCGCCGCAG 1849
     |::::::|||          :::|||:::::::::   :::   ::::::
 942 pAsnSerLys.........AsnLeuSerIleThrThrAsnSerSerSerT 956

1850 AAGACCGCACCCTGCTGCTTTCCGGCGGAACAAATTTAAACGGCAACATC 1899
          ||||||::::::   ::::::    ||||||   ||||||::::::
 956 hrTyrArgThrIleIleSerGlyAsnIleThrAsnLysAsnGlyAspLeu 972

1900 ACGCAAACAAACGGCAAACTGTTTTTCAGCGGCAGACCAACACCGCACGC 1949
     :::   ||||||                  :::|||      |||
 973 AsnIleThrAsn...............GluGlySerAspThrGlu..... 982

1950 CTACAATCATTTAAACGACCATTGGTCGCAAAAAGAGGGCATTCCTCGCG 1999
               :::::::::            |||||||||||          |
 983 ....MetGlnIleGlyGlyAspValSerGlnLysGlu...........G 994

2000 GGGAAATCGTGTGGGACAACGACTGGATCAACCGCACATTTAAAGCGGAA 2049
     ||:::::::::   ::::::|||   ||||||   |||            :::
 994 lyAsnLeuThrIleSerSerAspLysIleAsnIleThr.........Lys 1007

2050 AACTTCCAAATTAAAGGCGGACAGGCGGTGGTTTCCCGCAATGTTGCCAA 2099
     ::::::   ||||||:::|||                     |||
1008 GlnIleThrIleLysAlaGly....................ValAspGl 1017

2100 AGTGAAAGGCGATTGGCATTTGAGCAATCACGCCCAAGCAGTTTTTGGTG 2149
        ::::::|||         :::|||:::|||:::
1017 yGluAsnSerAspSerAspAlaThrAsnAsnAlaAsnLeu.......... 1030

2150 TCGCACCGCATCAAAGCCACACAATCTGTACACGTTCGGACTGGACGGGT 2199
                          ||||||   |||:::
1031 ....................ThrIleLysThrLys............... 1035

2200 CTGACAAATTGTGTCGAAAAAACCATTACCGACGATAAAGTGATTGCTTC 2249
                      |||      :::|||:::|||    |||:::::
1036 ...............GluLeuLysLeuThrGlnAspLeuAsnIleSerGl 1047

2250 ATTGACTAAGACCGACATCAGCGGCAATGTCGATCTTGCCGATCACGCTC 2299
     ::::::|||::::::|||::::::::   |||   :::|||
1047 yPheAsnLysAlaGluIleThrAlaLys...AspGlySerAsp....... 1060

2300 ATTTAAATCTCACAGGGCTTGCCACACTCAACGGCAATCTTAGTGCAAAT 2349
                          |||::::::::::::   ||||||:::
1061 ....................LeuThrIleGlyAsnThrAsnSerAlaAsp 1070

2350 GGCGATACACGTTATACAGTCAGCCACAACGCCACCCAAAACGGCAACCT 2399
     |||   :::          |||:::   |||   :::::::::::::::::
1071 GlyThrAsnAlaLysLysValThrPheAsnGlnValLysAspSerLysIl 1087

2400 TAGCCTCGTGGGCAATGCCCAAGCAACATTTAATCAAGCCACATTAAACG 2449
     :|||      |||                    :::::::::|||||:::::
1087 eSerAlaAspGly.................HisLysValThrLeuHisS 1098

2450 GCAACACATCGGCTTCGGGCAATGCTTCATTTAATCTAAGCGACCACGCC 2499
     ::::::::::::::||||||:::      :::   |||   :::|||    :::
1098 erLysValGluThrSerGlySerAsnAsn...AsnThrGluAspSerSer 1113

2500 GTACAAAACGGCAGTCTGACGCTTTCCGGCAACGCTAAGGCAAACGTAAG 2549
        :::|||:::::::||||||:::      :::||||||   ||||||::
1114 AspAsnAsnAlaGlyLeuThrIle......AspAlaLys...AsnValTh 1127

2550 CCATTCCGCACTCAACGGTAATGTCTCCCTAGCCGATAAGGCAGTATTCC 2599
     :          :::|||:::|||::::::   :::   |||||||||
1127 r.........ValAsnAsnAsnIleThr...SerHisLysAlaVal.... 1138

2600 ATTTGAAAGCAGCCGCTTTACCGGACAAATCAGCGGCGGCAAGGATACG 2649
          :::   |||      :::|||:::|||:::              |||
1139 .....SerIleSerAlaThrSerGlyGluIleThrThrLysThrGlyThr 1153

2650 GCATTACACTTAAAA......GACAGCGAATGGACGCTGCCGTCAGGCAC 2693
     ::::::::          :::   |||   |||      :::|||::
1154 ThrIleAsnAlaThrThrGlyAsnValGluIleThrAlaGlnThrGlySe 1170

2694 GGAATTAGGCAATTTAAACCTTGACAACGCCACCATTACACTCAATTCCG 2743
     :    ||||||:::::::::   ::::::::::::::||||||::::::::
1170 rIleLeuGlyGlyIleGluSerSerSerGlySerValThrLeuThrAlaT 1187

2744 CCTATCGCCACGATGCGGCAGGGGCGCAAACCGGCAGTGCGACAGATGCG 2793
     ::                |||::::::      :::|||:::::::::   :::
1187 hrGluGlyAlaLeuAlaValSerAsnIleSerGlyAsnThrValThrVal 1203

2794 CCGCGCCGCCGTTCGCGCCGTTCGCGCCGTTCCCTATTATCCGTTACACC 2843
         :::   :::        :::        :::        ::::::
1204 ThrAlaAsnSerGlyAlaLeuThrThrLeuAlaGlySerThrIleLysGl 1220

2844 GCCAACTTCGGTAGAATCCCGTTTCAACACGCTGACGGTAAACGGCAAAT 2893
        ||||||   :::         ::::::         :::::::|||   :
```

```
1220 yThrGluSerValThrThrSerSerGlnSerGlyAspIleGlyGlyThrI 1237

2894 TGAACGGTCAGGGAACA......TTCCGCTTTATGTCGGAACTCTTCGGC 2937
      ::::|||   |||||          :::      :::::|||
1237 leSerGly...GlyThrValGluValLysAlaThrGluSerLeuThrThr 1252

2938 TACCGCAGCGACAAATTGAAGCTGGCGGAAAGTTCCGAAGGCACTTACAC 2987
             ::::::|||:::|||   |||   ::::::|||:::::   ||
1253 GlnSerAsnSerLysIleLys...AlaThrThrGlyGluAlaAsnValTh 1268

2988 CTTGGCGGTC.......................AACAATACCGGCAACG 3013
      |   |||:::                      :::||||||   |||
1268 rSerAlaThrGlyThrIleGlyGlyThrIleSerGlyAsnThrValAsnV 1285

3014 AACCTGCAAGCCTCGAACAATTGACGGTA..................... 3042
          |||:::       :::|||||||||
1285 alThrAlaAsnAlaGlyAspLeuThrValGlyAsnGlyAlaGluIleAsn 1301

3043 ...GTGGAAGGAAAAGACAACAAACCGCTGTCCGAAAACCTTAATTTCAC 3089
        :::||||||       :::          |||::::::        ||
1302 AlaThrGluGlyAlaAlaThrLeuThrThrSerSerGlyLysLeuThrTh 1318

3090 CCTGCAAAACGAACACGTCGATGCCGGCGCGTGGCGTTACCAACTCATCC 3139
      |       ::::::|||:::   :::  |||:::|||    :::|||   |||
1318 rGluAlaSerSerHisIleThrSer.AlaLysGlyGlnValAsnLeuSer 1334

3140 GCAAAGACGGCGAGTTCCGCCTGCATAATCCGGTCAAAGAACAAGAGCTT 3189
      |||:::    :::|||   |||     |||::::::   |||    :::::
1335 AlaGlnAspGlySerValAlaGlySerIleAsnAlaAlaAsnValThrLe 1351

3190 TCCGACAAACTCGGCAAGGCAGAAGCCAAAAAACAGGCGGAAAAAGACAA 3239
      :     |||:::::::::                 |||:::   :::   ::::::|
1351 uAsnThrThrGlyThrLeuThrThrValLysGlySerAsnIleAsnAlaT 1368

3240 CGCGCAAAGCCTTG 3253
      ||        :::|||
1368 hrSerGlyThrLeu 1372
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01825

seq_documentation_block:

```
ID   AAB01825 standard; Protein; 1221 AA.
XX
AC   AAB01825;
XX
DT   11-SEP-2000  (first entry)
XX
DE   Haemophilus influenzae strain Joyc mature HMW1A protein, SEQ ID NO:28.
XX
KW   Mature HMW protein; hmw gene; hmwA1; hmwA2; high molecular weight;
KW   non-typeable Haemophilus influenzae; NTHi; non-encapsulated;
KW   recombinant production; Escherichia coli; antibacterial; vaccine;
KW   human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;
KW   detection; diagnosis.
XX
OS   Haemophilus influenzae strain Joyc.
XX
PN   WO200020609-A2.
XX
PD   13-APR-2000.
XX
PF   07-OCT-1999;   99WO-CA00938.
XX
PR   07-OCT-1998;   98US-0167568.
PR   08-DEC-1998;   98US-0206942.
XX
PA   (CONN-) CONNAUGHT LAB LTD.
XX
PI   Loosmore SM,  Yang Y,  Klein MH;
XX
DR   WPI; 2000-303789/26.
DR   N-PSDB; AAA52176.
XX
PT   Nucleic acid molecule for producing recombinant high molecular weight
PT   proteins of Haemophilus which are used as a vaccine to provide
PT   protection against Haemophilus induced diseases in humans -
XX
PS   Claim 8; Fig 18A-R; 307pp; English.
XX
CC   The invention relates to the recombinant production of Haemophilus
CC   influenzae high molecular weight (HMW) proteins in Escherichia coli. The
CC   expression construct used to effect recombinant expression comprises a
CC   promoter functional in E. coli (e.g., the T7 promoter) operably linked
CC   to a modified hmwABC operon from a non-typeable (non-encapsulated) H.
CC   influenzae (NTHi). Most HMW-expressing NTHi strains contain two hmw gene
CC   clusters termed hmw1ABC and hmw2ABC. Each hmwABC operon comprises hmwA,
CC   hmwB and hmwC genes. The hmwA genes encode the structural HMWA proteins
CC   and the hmwB and hmwC genes encode accessory proteins which are
CC   responsible for post-translational processing and secretion of the HMWA
CC   proteins. The modified hmwABC operon used in the expression construct of
CC   the invention contains an A gene modified such that it encodes only the
CC   mature HMWA. The invention also discloses hmwA genes (AAA52175-A52198)
CC   and HMWA proteins (AAB01824-B01849) from the non-typeable H. influenzae
CC   strains Joyc, K1, K21, LCDC2, PMH1, 15 and 12. The nucleic acids and
CC   vectors are used for the production of recombinant H. influenzae HMW
CC   proteins which can be used as vaccines to mediate a humoral or
CC   cell-mediated immune response to provide protection against diseases in
CC   humans caused by H. influenzae (e.g., otitis media, epiglottitis,
CC   pneumonia and tracheobronchitis). The HMW proteins are also useful as
CC   antigens in immunoassays for detecting antibodies against Haemophilus,
CC   HMW proteins and/or HMW peptides. The nucleotide sequences encoding the
CC   HMW proteins can be used to isolate and clone hmw genes from other
CC   non-typeable strains of Haemophilus via hybridisation reactions. The
CC   present sequence represents a mature HMWA protein from a non-typeable
CC   strain of H. influenzae.
XX
SQ   Sequence   1221 AA;
```

alignment_scores:

```
            Quality:  265.00          Length:    900
              Ratio:  0.559             Gaps:     48
 Percent Similarity:  52.667  Percent Identity:  21.222
```

alignment_block:

US-09-303-518D-649 x AAB01825 ..

Align seg 1/1 to: AAB01825 from: 1 to: 1221

```
712 TCAGGTGGTGGCACAGTCAACTTAGGTAGTGAAAAA.........ATTAA 752
     :::|||   ||||||::::::   |||:::   |||          :::::
146 ThrGlyGlnGlyThrIleThrAlaGlyAsnGlyLysGlyPheArgPheGl 162

753 ACATAGCCCATATGGTTTTTTACCAACAGGAGGCTCATTTGGC....... 795
     ::::::       :::   :::   ||||||     |||:::
162 uAsnAlaSerLeuAsnGlyIleGlyThrGlyLeuLeuPheAsnIleLysA 179

796 ..GACAGTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCAAAAGTGG 843
       |||   |||:::   :::   |||   :::
179 rgAspLeuGlyAsnAsnPheGlnIleIleAsn.................. 189

844 TTAATTAATGGGGTATTGCAAACGGGCAACCCCTATATAGGAAAAAGCAA 893
     ::::::||||||:::|||:::   :::             ||||||   ||
190 PhePheAsnGlyThrLeuAsnIleSer...........GlyLysValAs 202

894 TGGCTTCCAGCTGGTTCGTAAAGATTGGTTCTATGATGAAATCTTTGCTG 943
     |      :::::::::   |||   |||   |||::::::   |||   |
202 nIleSerMetValIleProLysLysTrpAspTyrSerLys...PheArgG 218

944 GAGATACCCAT......TCAGTATTCTACGAACCACGTCAAAATGGGAAA 987
     ||   |||:::          :::       :::      :::::::::|||
218 lyArgThrTyrTrpAsnValThrHisLeuAsnValSerGluGlySerLys 234

988 TACTCTTTTAACGACGATAATAATGGC.........ACAGGAAAAATCAA 1028
     ::::::::::::   |||::::::|||          :::|||   :::||
235 PheAsnLeuThrIleAspSerArgGlyAspAspThrAlaGlyThrLeuAs 251
```

```
1029 TGCCAAACATGAACACAATTCTCTGCCTAATAGATTAAAAACACGAACCG 1078
     |:::    ::::::   |||::::::        :::       |||
 251 nThrProTyrAsnLeuAsnGlyIleSerPheAsnLysAspThr....... 265

1079 TTCAATTGTTTAATGTT...........TCTTTATCCGAGACAGCAAGA 1116
          :::|||:::|||            :::::::::          :::
 266 .....IlePheAspValLysGlnAsnGlyAlaValThrPheAspIleLys 280

1117 GAACCTGTTTATCATGCTGCAGGTGGTGTCAACAGTTATCGACCC..... 1161
        |||:::                |||||||||:::    |||
 281 AlaProIle..............GlyValAsnAsnAsnArgAsnLeuAs 292

1162 .........................AGACTGAATAATGGAGAAAATATTT 1186
                              :::::::::|||    |||::::
 292 nTyrAlaSerPheAsnGlyAsnIleSerValSerGlyGlyGlyAsnValT 309

1187 CCTTT.....................ATTGACGAAGGAAAAGGCGAATTG 1215
     ::|||                         :::       |||    :::
 309 hrPheLysLeuLeuAlaSerSerSerThrAlaGlnThrProGlyValPhe 325

1216 ATACTTACCAGCAACATCAAT...CAAGGTGCTGGAGGATTATATTTCCA 1262
     |||     ::::::::::::|||     :::|||::::::::::|||     |||::
 326 IleAsnSerLysHisPheAsnAlaSerGlyGlySerSerLeuGluPheAr 342

1263 AGGAGAT.................TTTACGGTCTCGCCTGAAAATAACG 1294
     :    :::                     |||    ::::::      |||:::
 342 gThrGluGlySerThrLysValGlyPheLeuIleAsn......AsnAspL 357

1295 AAACTTGGCAAGGCGCGGGCGTTCATATCAGTGAAGACAGTACCGTTACT 1344
        |||    :::::::::|||     :::||||||
 357 euThrLeuAsnAlaThrGlyGlyAsnIleSer..............Leu 368

1345 TGGAAAGTAAACGGCGTGGCAAACGACCGCCTGTCCAAAATCGGCAAAGG 1394
        :::|||:::|||:::                 :::    |||||||||||
 369 LeuGlnValGluGlyIleAsp............GlyMetIleGlyLysGl 381

1395 CACGCTGCACGTTCAAGCCAAAGGGGAAAACCAAGGCTCGATCAGCGTGG 1444
     |          |||    ||||||                :::|||:::    :
 381 y.........ValValAlaLys...........LysAsnIleThrPheA 391

1445 GCGACGGTACAGTCATTTTGGATCAGCAGGCAGACGATAAAGGCAAAAAA 1494
     ::    |||::::::    :::                        |||:::|||
 391 laGlyGlyAsnIleThrPhe....................GlySerLys 400

1495 CAAGCCTTTAGTGAAATCGGCTTGGTCAGCGGCAGGGGTACGGTGCAACT 1544
     :::|||::::::||||||               :::|||:::::::|||::::::
 401 LysAlaIleThrGluIle.........GluGlyAsnAlaThrIleAsnAs 414

1545 GAATGCC..........................GATAATCAGTTCAACC 1567
        ||||||                               ||||||:::    :::|
 414 nAsnAlaAsnValThrLeuIleGlySerAspPheAspAsnHisGlnLysP 431

1568 CC..............GACAAACTCTATTTCGGCTTTCGCGGCGGACGT 1602
     ||                   |||     :::            :::::::::|||:::
 431 roLeuThrIleLysLysAspValIle.........IleAsnSerGlyAsn 444

1603 TTGGATTTAAACGGGCATTCGCTTTCGTTCCACCGTATTCAAAATACCGA 1652
     |||        :::|||:::     ::::::::::::
 445 LeuThrAlaGlyGlyAsnValIleAsnIleAsn................ 455

1653 TGAAGGGGCGATGATTGTCAACCACAATCAAGACAAAGAATCCACCGTTA 1702
          |||     :::     |||||:::::     :::     ::::::     ::::
 456 ....GlyAsnLeuThrValAsnAsnGlyAlaAsnLeuLysAlaIleThrA 471

1703 CCATTACAGGCAATAAAGATATTGCTACAACCGGCAAT......AACAAC 1746
     ::::::|||    |||                        :::|||:::      :::|||
 471 snPheThrPheAsn..............ValGlyGlyLeuPheAspAsn 482

1747 AGCTTGGATAGCAAAAAAGAAATTGCCTACAACGGTTGGTTTGGCGAGAA 1796
     :::    :::|||:::    :::||||||    :::|||              ||
 483 LysGlyAsnSerAsnIleSerIleAla...ArgGlyGlyAlaLysPheLy 498

1797 AGATACGACCAAAACGAACGGGCGGCTCAACCTTGTTTACCAGCCCGCCG 1846
     ||||    ::::::|||::::::     ||||||::::::    :::        :
 498 sAspIleAsnAsnThrSerSer...LeuAsnIleThrThrAsnSerAspT 514

1847 CAGAAGACCGCACCCTGCTGCTTTCCGGCGGAACAAATTTAAACGGCAAC 1896
     ::        ||||||::::::     ::::::     ||||||      |||:::
 514 hrThrTyrArgThrIleIleGluGlyAsnIleThrAsnLysAlaGlyAsp 530

1897 ATCACGCAAACAAACGGCAAACTGTTTTTCAGCGGCAGACCAACACCGCA 1946
     ::::::        :::::|||                            ::
 531 LeuAsnIleIleAspAsnLys......................GlyAs 539

1947 CGCCTACAATCATTTAAACGACCATTGGTCGCAAAAAGAGGGCATTCCTC 1996
     :|||        :::::::::    :::    ||||||||||||
 539 nAlaGluIleGlnIleGlyGlyAsnIleSerGlnLysGlu.......... 552

1997 GCGGGGAAATCGTGTGGGACAACGACTGGATCAACCGCACATTTAAAGCG 2046
       |||:::::::::    ::::::|||    ||||||    |||
 553 ..GlyAsnLeuThrIleSerSerAspLysIleAsnIleThr......... 565

2047 GAAAACTTCCAAATTAAAGGCGGA.......................... 2070
     :::::::::    ||||||    |||
 566 AsnGlnIleThrIleLysLysGlyValAsnLysGluAspSerAspSerSe 582

2071 ................CAGGCGGTGGTTTCCCGCAATGTTGCCAAAGTGA 2104
                     :::    :::::::::    :::       ::::::
 582 rThrAlaAsnAsnAlaAsnLeuThrIleLysThrLysGluLeuGlnLeuT 599

2105 AAGGCGATTGGCATTTGAGCAATCACGCCCAAGCAGTTTTGGTGTCGCA 2154
       ||||||    ::::::|||:::          :::|||    :::    :::|||
 599 hrGlyAspLeuAsnIleSerGlyPheAspLysAlaGluIle...ThrAla 614

2155 CCGCATCAAAGCCACACAATCTGTACACGTTCGGACTGGACGGGTCTGAC 2204
       :::    :::          |||       :::||||||    ::::::    ::
 615 LysGluGlyAlaAspLeuIleIleGlyAsnSerAspAsnAsnAsnAl 631

2205 AAATTGTGTCGAAAAAAACC.........ATTACCGACGATAAAGTGATTG 2245
     :|||:::    :::    |||                :::    |||:::|||:::    |
 631 aAsnAlaLysLysValThrPheAsnGlnValLysAspSerLysIleSerA 648

2246 CTTCATTGACTAAGACCGACATCAGCGGCAATGTCGATCTTGCC...... 2289
     |||:::    ::::::    ::::::::::::|||:::    :::
 648 laGlySerHisAsnValThrLeuAsnSerLysValGluThrSerAsnGly 664

2290 ..........................GATCACGCTCATTTAAATCTCAC 2312
                                    |||    :::    |||::::::::
 665 AsnAsnAspAlaGluSerAsnAsnGlyAspSerThrSerLeuThrIleAs 681

2313 AGGGCTT...GCCACACTCAACGGCAATCTTAGTGCAAATGGCGATACAC 2359
     ::::       :::|||:::|||:::|||:::::::::::::             ::::
 681 nAlaLysAsnValThrValAsnAsnAsnIleThrSerHisLysThrValA 698

2360 GTTATACAGTCAGCCACAACGCCACCCAAAACGGCAACCTTAGCCTC... 2406
     ::      |||:::|||:::|||:::|||    :::::::::    ::::::
 698 snIleThrAlaSerGluAsnValThrThrLysAlaGlyThrThrIleAsn 714

2407 ......GTGGGCAATGCCCAAGCAACATTTAATCAAGCCACATTAAACGG 2450
           :::|||::::::::::::|||    :::    :::    ::::::||
 715 AlaThrIleGlySerValGluValThrAlaLysThrGlyAspIleLysGl 731

2451 C..........................AACACATCGGCTTCGGGCAATG 2473
     |                              |||    :::||||||||||:::
 731 yGlyIleGluSerAsnSerGlyAsnValAsnIleThrAlaSerGlyAsp. 747

2474 CTTCATTTAATCTAAGCGACCACGCCGTACAAAACGGCAGTCTGACGCTT 2523
       ::::::|||:::|||:::    :::    ||||||    :::::::::
 748 ..ThrLeuAsnValSerAsnIleThrGlyGlnAsnValThrValAlaAla 763

2524 TCCGGCAACGCTAAGGCAAACGTAAGCCATTCCGCACTCAAC........ 2565
     :::::::::|||    ::::::::::::    |||::::::|||
 764 AlaSerGlyAlaValThrThrThrLysGlySerThrIleAsnAlaThrTh 780

2566 .GGTAATGTCTCCCTAGCCGATAAGGCAGTATTCCATTTGAAAGCAGCC 2614
```

```
         ||||||:::::::::::   |||
 780 rGlyAsnAlaAsnIleThrThrLys......................... 788

2615 GCTTTACCGGACAAATCAGCGGCGGCAAGGATACGGCA.........TTA 2655
         ||||||:::|||:::|||          :::|||          :::
 789 .....ThrGlyGluIleAsnGlyGluValLysSerAlaSerGlyAsnVal 803

2656 CACTTAAAAGACAGCGAATGGACGCTGCCGTCAGGCACGGAATTAGGCAA 2705
     ::::::      |||     ||||||         :::::::::::||
 804 AsnIleThrAlaSerGlyAsnThrLeu.........AsnAsnValSerAs 817

2706 TTTAAACCTTGACAACGCCACCATTACACTCAATTCCGCCTATCGCCACG 2755
     |::::::   :::|||:::|||:::|||    |||||||:::
 817 nIleThrGlyGlnAsnValThrValThrAlaAsnSerGlyAlaIleThrT 834

2756 ATGCGGCAGGGGCGCAAACCGGCAGTGCGACA...GATGCGCCGCGCCGC 2802
       :::   |||:::      :::::::::|||    ||||||
 834 hrThrGluGlySerThrIleAsnAlaThrThrGlyAspAla......... 847

2803 CGTTCGCGCCGTTCGCGCCGTTCCCTATTATCCGTTACACCGCCAACTTC 2852

 847 .................................................. 847

2853 GGTAGAATCCCGTTTCAACACGCTGACGGTAAACGGCAAATTGAACGGTC 2902
                      |||      |||    :::|||::::::||||||:
 848 ................AsnIleThrThrGlnThrGlyAsnIleAsnGlyL 859

2903 AGGGAACATTCCGCTTTATGTCGGAACTCTTCGGCTACCGCAGCGACAAA 2952
     ::
 859 ys................................................ 859

2953 TTGAAGCTGGCGGAAAGTTCCGAAGGCACTTACACCTTGGCGGTCAACAA 3002
              :::|||||||||:::|||:::    ||||||   ::::::::
 860 .........ValGluSerSerSerGlySerValThrLeuIleAlaThrGl 873

3003 T...........ACCGGCAACGAACCTGCAAGCCTCGAACAATTGACGG 3040
     :           :::||||||        ::::::        :::|||:
 873 yGlnThrLeuAlaValGlyAsnIleSerGlyAspThrValThrIleThrA 890

3041 TAGTGGAAGGAAAAGACAACAAACCGCTGTCCGAAAACCTTAATTTCACC 3090
     ::   :::||||||   :::             |||::::::::|||   |||
 890 laAspLysGlyLysLeuThrThrGlnThrSerSerLysIleAsnGlyThr 906
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01824

seq_documentation_block:

```
ID   AAB01824 standard; Protein; 1227 AA.
XX
AC   AAB01824;
XX
DT   11-SEP-2000  (first entry)
XX
DE   Haemophilus influenzae strain Joyc HMW1A protein, SEQ ID NO:26.
XX
KW   HMW protein; hmw gene; hmwA1; hmwA2; high molecular weight;
KW   non-typeable Haemophilus influenzae; NTHi; non-encapsulated;
KW   recombinant production; Escherichia coli; antibacterial; vaccine;
KW   human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;
KW   detection; diagnosis.
XX
OS   Haemophilus influenzae strain Joyc.
XX
PN   WO200020609-A2.
XX
PD   13-APR-2000.
XX
PF   07-OCT-1999;    99WO-CA00938.
XX
PR   07-OCT-1998;    98US-0167568.
PR   08-DEC-1998;    98US-0206942.
XX
PA   (CONN-) CONNAUGHT LAB LTD.
XX
PI   Loosmore SM,  Yang Y,  Klein MH;
XX
DR   WPI; 2000-303789/26.
DR   N-PSDB; AAA52175.
XX
PT   Nucleic acid molecule for producing recombinant high molecular weight
PT   proteins of Haemophilus which are used as a vaccine to provide
PT   protection against Haemophilus induced diseases in humans -
XX
PS   Claim 12; Fig 18A-R; 307pp; English.
XX
CC   The invention relates to the recombinant production of Haemophilus
CC   influenzae high molecular weight (HMW) proteins in Escherichia coli. The
CC   expression construct used to effect recombinant expression comprises a
CC   promoter functional in E. coli (e.g., the T7 promoter) operably linked
CC   to a modified hmwABC operon from a non-typeable (non-encapsulated) H.
CC   influenzae (NTHi). Most HMW-expressing NTHi strains contain two hmw gene
CC   clusters termed hmw1ABC and hmw2ABC. Each hmwABC operon comprises hmwA,
CC   hmwB and hmwC genes. The hmwA genes encode the structural HMWA proteins
CC   and the hmwB and hmwC genes encode accessory proteins which are
CC   responsible for post-translational processing and secretion of the HMWA
CC   proteins. The modified hmwABC operon used in the expression construct of
CC   the invention contains an A gene modified such that it encodes only the
CC   mature HMWA. The invention also discloses hmwA genes (AAA52175-A52198)
CC   and HMWA proteins (AAB01824-B01849) from the non-typeable H. influenzae
CC   strains Joyc, K1, K21, LCDC2, PMH1, 15 and 12. The nucleic acids and
CC   vectors are used for the production of recombinant H. influenzae HMW
CC   proteins which can be used as vaccines to mediate a humoral or
CC   cell-mediated immune response to provide protection against diseases in
CC   humans caused by H. influenzae (e.g., otitis media, epiglottitis,
CC   pneumonia and tracheobronchitis). The HMW proteins are also useful as
CC   antigens in immunoassays for detecting antibodies against Haemophilus,
CC   HMW proteins and/or HMW peptides. The nucleotide sequences encoding the
CC   HMW proteins can be used to isolate and clone hmw genes from other
CC   non-typeable strains of Haemophilus via hybridisation reactions. The
CC   present sequence represents an HMWA protein from a non-typeable strain of
CC   H. influenzae.
XX
SQ   Sequence   1227 AA;
```

alignment_scores:

```
           Quality:  265.00             Length:    900
             Ratio:  0.559                Gaps:     48
Percent Similarity:  52.667   Percent Identity:  21.222
```

alignment_block:

```
 US-09-303-518D-649 x AAB01824   ..

 Align seg 1/1  to: AAB01824  from: 1  to: 1227

 712 TCAGGTGGTGGCACAGTCAACTTAGGTAGTGAAAAA.........ATTAA 752
     :::|||   ||||||::::::   |||:::   |||          :::::
 152 ThrGlyGlnGlyThrIleThrAlaGlyAsnGlyLysGlyPheArgPheGl 168

 753 ACATAGCCCATATGGTTTTTTACCAACAGGAGGCTCATTTGGC....... 795
     :::::::         :::   :::   ||||||      |||:::
 168 uAsnAlaSerLeuAsnGlyIleGlyThrGlyLeuLeuPheAsnIleLysA 185

 796 ..GACAGTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCAAAAGTGG 843
       |||   |||:::   :::   |||   :::
 185 rgAspLeuGlyAsnAsnPheGlnIleIleAsn.................. 195

 844 TTAATTAATGGGGTATTGCAAACGGGCAACCCCTATATAGGAAAAAGCAA 893
     ::::::||||||:::|||:::   :::                ||||||   ||
 196 PhePheAsnGlyThrLeuAsnIleSer............GlyLysValAs 208

 894 TGGCTTCCAGCTGGTTCGTAAAGATTGGTTCTATGATGAAATCTTTGCTG 943
     |        :::::::::   |||   |||   |||::::::   |||   |
 208 nIleSerMetValIleProLysLysTrpAspTyrSerLys...PheArgG 224

 944 GAGATACCCAT......TCAGTATTCTACGAACCACGTCAAAATGGGAAA 987
     ||   |||:::           :::       :::       :::::::::|||
```

```
 224 lyArgThrTyrTrpAsnValThrHisLeuAsnValSerGluGlySerLys 240

 988 TACTCTTTTAACGACGATAATAATGGC.........ACAGGAAAAATCAA 1028
     :::::::::::   |||:::::|||          :::|||   ::::||
 241 PheAsnLeuThrIleAspSerArgGlyAspAspThrAlaGlyThrLeuAs 257

1029 TGCCAAACATGAACACAATTCTCTGCCTAATAGATTAAAAACACGAACCG 1078
     |:::   ::::::   |||:::::        :::       |||
 257 nThrProTyrAsnLeuAsnGlyIleSerPheAsnLysAspThr....... 271

1079 TTCAATTGTTTAATGTT...........TCTTTATCCGAGACAGCAAGA 1116
          :::|||:::|||          :::::::::         ::::
 272 .....IlePheAspValLysGlnAsnGlyAlaValThrPheAspIleLys 286

1117 GAACCTGTTTATCATGCTGCAGGTGGTGTCAACAGTTATCGACCC..... 1161
       |||:::                 ||||||||:::    |||
 287 AlaProIle...............GlyValAsnAsnAsnArgAsnLeuAs 298

1162 ........................AGACTGAATAATGGAGAAAATATTT 1186
                                :::::::|||   |||::::
 298 nTyrAlaSerPheAsnGlyAsnIleSerValSerGlyGlyGlyAsnValT 315

1187 CCTTT....................ATTGACGAAGGAAAAGGCGAATTG 1215
     ::|||                        :::        |||    :::
 315 hrPheLysLeuLeuAlaSerSerSerThrAlaGlnThrProGlyValPhe 331

1216 ATACTTACCAGCAACATCAAT...CAAGGTGCTGGAGGATTATATTTCCA 1262
     |||    ::::::::::|||    :::|||:::::::::|||    |||::
 332 IleAsnSerLysHisPheAsnAlaSerGlyGlySerSerLeuGluPheAr 348

1263 AGGAGAT.................TTTACGGTCTCGCCTGAAAATAACG 1294
     :   :::                     |||    ::::::      |||:::
 348 gThrGluGlySerThrLysValGlyPheLeuIleAsn......AsnAspL 363

1295 AAACTTGGCAAGGCGCGGGCGTTCATATCAGTGAAGACAGTACCGTTACT 1344
        |||    :::::::::|||    :::||||||
 363 euThrLeuAsnAlaThrGlyGlyAsnIleSer..............Leu 374

1345 TGGAAAGTAAACGGCGTGGCAAACGACCGCCTGTCCAAAATCGGCAAAGG 1394
        :::|||:::|||:::                  :::    |||||||||||
 375 LeuGlnValGluGlyIleAsp...........GlyMetIleGlyLysGl 387

1395 CACGCTGCACGTTCAAGCCAAAGGGGAAAACCAAGGCTCGATCAGCGTGG 1444
     |           |||    ||||||                :::|||:::    :
 387 y.........ValValAlaLys...........LysAsnIleThrPheA 397

1445 GCGACGGTACAGTCATTTGGATCAGCAGGCAGACGATAAAGGCAAAAAA 1494
     ::    |||:::::    :::                    |||:::|||
 397 laGlyGlyAsnIleThrPhe....................GlySerLys 406

1495 CAAGCCTTTAGTGAAATCGGCTTGGTCAGCGGCAGGGGTACGGTGCAACT 1544
     :::|||:::::::||||||         :::|||:::::::|||::::::
 407 LysAlaIleThrGluIle.........GluGlyAsnAlaThrIleAsnAs 420

1545 GAATGCC..........................GATAATCAGTTCAACC 1567
       ||||||                            |||||||:::    :::|
 420 nAsnAlaAsnValThrLeuIleGlySerAspPheAspAsnHisGlnLysP 437

1568 CC..............GACAAACTCTATTTCGGCTTTCGCGGCGGACGT 1602
      ||                    |||    :::           ::::::::|||:::
 437 roLeuThrIleLysLysAspValIle.........IleAsnSerGlyAsn 450

1603 TTGGATTTAAACGGGCATTCGCTTTCGTTCCACCGTATTCAAAATACCGA 1652
      |||        :::|||:::    ::::::::::::
 451 LeuThrAlaGlyGlyAsnValIleAsnIleAsn................ 461

1653 TGAAGGGGCGATGATTGTCAACCACAATCAAGACAAAGAATCCACCGTTA 1702
          |||    :::    ||||||::::::    :::    ::::::    ::::
 462 ....GlyAsnLeuThrValAsnAsnGlyAlaAsnLeuLysAlaIleThrA 477

1703 CCATTACAGGCAATAAAGATATTGCTACAACCGGCAAT......AACAAC 1746
     ::::::|||    |||                   :::|||:::      :::|||
 477 snPheThrPheAsn..............ValGlyGlyLeuPheAspAsn 488

1747 AGCTTGGATAGCAAAAAAGAAATTGCCTACAACGGTTGGTTTGGCGAGAA 1796
     :::    :::|||:::    :::||||||    :::|||            ||
 489 LysGlyAsnSerAsnIleSerIleAla...ArgGlyGlyAlaLysPheLy 504

1797 AGATACGACCAAAACGAACGGGCGGCTCAACCTTGTTTACCAGCCCGCCG 1846
     ||||    ::::::|||::::::    ||||||::::::    :::        :
 504 sAspIleAsnAsnThrSerSer...LeuAsnIleThrThrAsnSerAspT 520

1847 CAGAAGACCGCACCCTGCTGCTTTCCGGCGGAACAAATTTAAACGGCAAC 1896
     ::          ||||||::::::    ::::::    ||||||        |||:::
 520 hrThrTyrArgThrIleIleGluGlyAsnIleThrAsnLysAlaGlyAsp 536

1897 ATCACGCAAACAAACGGCAAACTGTTTTTCAGCGGCAGACCAACACCGCA 1946
     ::::::          :::::::|||                          ::
 537 LeuAsnIleIleAspAsnLys.......................GlyAs 545

1947 CGCCTACAATCATTTAAACGACCATTGGTCGCAAAAAGAGGGCATTCCTC 1996
     :|||        ::::::::    :::    ||||||||||||
 545 nAlaGluIleGlnIleGlyGlyAsnIleSerGlnLysGlu.......... 558

1997 GCGGGGAAATCGTGTGGGACAACGACTGGATCAACCGCACATTTAAAGCG 2046
        |||:::::::::    :::::::|||    ||||||    |||
 559 ..GlyAsnLeuThrIleSerSerAspLysIleAsnIleThr......... 571

2047 GAAAACTTCCAAATTAAAGGCGGA........................ 2070
     :::::::::        ||||||    |||
 572 AsnGlnIleThrIleLysLysGlyValAsnLysGluAspSerAspSerSe 588

2071 ................CAGGCGCTGGTTTCCCGCAATGTTGCCAAAGTGA 2104
                          :::    :::::::::    :::        ::::::
 588 rThrAlaAsnAsnAlaAsnLeuThrIleLysThrLysGluLeuGlnLeuT 605

2105 AAGGCGATTGGCATTTGAGCAATCACGCCCAAGCAGTTTTTGGTGTCGCA 2154
        ||||||    ::::::|||:::        :::|||    :::    :::|||
 605 hrGlyAspLeuAsnIleSerGlyPheAspLysAlaGluIle...ThrAla 620

2155 CCGCATCAAAGCCACACAATCTGTACACGTTCGGACTGGACGGGTCTGAC 2204
         :::    :::        |||        :::||||||    ::::::    ::
 621 LysGluGlyAlaAspLeuIleIleGlyAsnSerAspAsnAsnAsnAsnAl 637

2205 AAATTGTGTCGAAAAAAACC.........ATTACCGACGATAAAGTGATTG 2245
     :|||:::    :::    |||             :::    |||:::|||:::    |
 637 aAsnAlaLysLysValThrPheAsnGlnValLysAspSerLysIleSerA 654

2246 CTTCATTGACTAAGACCGACATCAGCGGCAATGTCGATCTTGCC...... 2289
     ||:::        ::::::    ::::::::::::|||:::    :::
 654 laGlySerHisAsnValThrLeuAsnSerLysValGluThrSerAsnGly 670

2290 ..........................GATCACGCTCATTTAAATCTCAC 2312
                                   |||    :::    |||::::::::
 671 AsnAsnAspAlaGluSerAsnAsnGlyAspSerThrSerLeuThrIleAs 687

2313 AGGGCTT...GCCACACTCAACGGCAATCTTAGTGCAAATGGCGATACAC 2359
     ::::          :::|||:::|||:::|||:::::::::::::        ::::
 687 nAlaLysAsnValThrValAsnAsnAsnIleThrSerHisLysThrValA 704

2360 GTTATACAGTCAGCCACAACGCCACCCAAAACGGCAACCTTAGCCTC... 2406
     ::    |||:::|||:::|||:::|||    :::::::::    ::::::
 704 snIleThrAlaSerGluAsnValThrThrLysAlaGlyThrThrIleAsn 720

2407 ......GTGGGCAATGCCCAAGCAACATTTAATCAAGCCACATTAAACGG 2450
            :::|||::::::::::::|||    :::    :::    ::::::||
 721 AlaThrIleGlySerValGluValThrAlaLysThrGlyAspIleLysGl 737

2451 C..........................AACACATCGGCTTCGGGCAATG 2473
     |                              |||    :::||||||||||:::
 737 yGlyIleGluSerAsnSerGlyAsnValAsnIleThrAlaSerGlyAsp. 753

2474 CTTCATTTAATCTAAGCGACCACGCCGTACAAAACGGCAGTCTGACGCTT 2523
        ::::::|||:::|||:::    :::    ||||||    :::::::::
 754 ..ThrLeuAsnValSerAsnIleThrGlyGlnAsnValThrValAlaAla 769
```

```
2524 TCCGGCAACGCTAAGGCAAACGTAAGCCATTCCGCACTCAAC........ 2565
     :::::::::|||   ::::::::::::   |||::::::|||
 770 AlaSerGlyAlaValThrThrThrLysGlySerThrIleAsnAlaThrTh 786

2566 .GGTAATGTCTCCCTAGCCGATAAGGCAGTATTCCATTTTGAAAGCAGCC 2614
      ||||||:::::::::::::   |||
 786 rGlyAsnAlaAsnIleThrThrLys......................... 794

2615 GCTTTACCGGACAAATCAGCGGCGGCAAGGATACGGCA.........TTA 2655
          ||||||:::|||:::|||           :::|||        :::
 795 .....ThrGlyGluIleAsnGlyGluValLysSerAlaSerGlyAsnVal 809

2656 CACTTAAAAGACAGCGAATGGACGCTGCCGTCAGGCACGGAATTAGGCAA 2705
     ::::::      |||      ||||||         ::::::::::::||
 810 AsnIleThrAlaSerGlyAsnThrLeu.........AsnAsnValSerAs 823

2706 TTTAAACCTTGACAACGCCACCATTACACTCAATTCCGCCTATCGCCACG 2755
     |::::::   :::|||:::|||:::|||   |||||||:::
 823 nIleThrGlyGlnAsnValThrValThrAlaAsnSerGlyAlaIleThrT 840

2756 ATGCGGCAGGGGCGCAAACCGGCAGTGCGACA...GATGCGCCGCGCCGC 2802
       :::    |||:::       :::::::::|||    ||||||
 840 hrThrGluGlySerThrIleAsnAlaThrThrGlyAspAla......... 853

2803 CGTTCGCGCCGTTCGCGCCGTTCCCTATTATCCGTTACACCGCCAACTTC 2852

 853 .................................................. 853

2853 GGTAGAATCCCGTTTCAACACGCTGACGGTAAACGGCAAATTGAACGGTC 2902
                        |||      |||    :::|||::::::||||||:
 854 ................AsnIleThrThrGlnThrGlyAsnIleAsnGlyL 865

2903 AGGGAACATTCCGCTTTATGTCGGAACTCTTCGGCTACCGCAGCGACAAA 2952
     ::
 865 ys................................................ 865

2953 TTGAAGCTGGCGGAAAGTTCCGAAGGCACTTACACCTTGGCGGTCAACAA 3002
              :::|||||||||:::|||:::    ||||||    ::::::::
 866 .........ValGluSerSerSerGlySerValThrLeuIleAlaThrGl 879

3003 T............ACCGGCAACGAACCTGCAAGCCTCGAACAATTGACGG 3040
     :            :::|||||||      ::::::        :::|||:
 879 yGlnThrLeuAlaValGlyAsnIleSerGlyAspThrValThrIleThrA 896

3041 TAGTGGAAGGAAAAGACAACAAACCGCTGTCCGAAAACCTTAATTTCACC 3090
     ::    :::||||||    :::          |||::::::::::|||    |||
 896 laAspLysGlyLysLeuThrThrGlnThrSerSerLysIleAsnGlyThr 912
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01835

seq_documentation_block:

```
ID   AAB01835 standard; Protein; 1095 AA.
XX
AC   AAB01835;
XX
DT   11-SEP-2000  (first entry)
XX
DE   Haemophilus influenzae strain LCDC2 mature HMW1A protein, SEQ ID NO:45.
XX
KW   Mature HMW protein; hmw gene; hmwA1; hmwA2; high molecular weight;
KW   non-typeable Haemophilus influenzae; NTHi; non-encapsulated;
KW   recombinant production; Escherichia coli; antibacterial; vaccine;
KW   human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;
KW   detection; diagnosis.
XX
OS   Haemophilus influenzae strain LCDC2.
XX
PN   WO200020609-A2.
XX
PD   13-APR-2000.
XX
PF   07-OCT-1999;   99WO-CA00938.
XX
PR   07-OCT-1998;   98US-0167568.
PR   08-DEC-1998;   98US-0206942.
XX
PA   (CONN-) CONNAUGHT LAB LTD.
XX
PI   Loosmore SM,  Yang Y,  Klein MH;
XX
DR   WPI; 2000-303789/26.
DR   N-PSDB; AAA52184.
XX
PT   Nucleic acid molecule for producing recombinant high molecular weight
PT   proteins of Haemophilus which are used as a vaccine to provide
PT   protection against Haemophilus induced diseases in humans -
XX
PS   Claim 8; Fig 22A-P; 307pp; English.
XX
CC   The invention relates to the recombinant production of Haemophilus
CC   influenzae high molecular weight (HMW) proteins in Escherichia coli. The
CC   expression construct used to effect recombinant expression comprises a
CC   promoter functional in E. coli (e.g., the T7 promoter) operably linked
CC   to a modified hmwABC operon from a non-typeable (non-encapsulated) H.
CC   influenzae (NTHi). Most HMW-expressing NTHi strains contain two hmw gene
CC   clusters termed hmw1ABC and hmw2ABC. Each hmwABC operon comprises hmwA,
CC   hmwB and hmwC genes. The hmwA genes encode the structural HMWA proteins
CC   and the hmwB and hmwC genes encode accessory proteins which are
CC   responsible for post-translational processing and secretion of the HMWA
CC   proteins. The modified hmwABC operon used in the expression construct of
CC   the invention contains an A gene modified such that it encodes only the
CC   mature HMWA. The invention also discloses hmwA genes (AAA52175-A52198)
CC   and HMWA proteins (AAB01824-B01849) from the non-typeable H. influenzae
CC   strains Joyc, K1, K21, LCDC2, PMH1, 15 and 12. The nucleic acids and
CC   vectors are used for the production of recombinant H. influenzae HMW
CC   proteins which can be used as vaccines to mediate a humoral or
CC   cell-mediated immune response to provide protection against diseases in
CC   humans caused by H. influenzae (e.g., otitis media, epiglottitis,
CC   pneumonia and tracheobronchitis). The HMW proteins are also useful as
CC   antigens in immunoassays for detecting antibodies against Haemophilus,
CC   HMW proteins and/or HMW peptides. The nucleotide sequences encoding the
CC   HMW proteins can be used to isolate and clone hmw genes from other
CC   non-typeable strains of Haemophilus via hybridisation reactions. The
CC   present sequence represents a mature HMWA protein from a non-typeable
CC   strain of H. influenzae.
XX
SQ   Sequence   1095 AA;
```

alignment_scores:

```
           Quality:  264.50            Length:   1153
             Ratio:  0.442               Gaps:     56
 Percent Similarity:  51.865   Percent Identity: 20.035
```

alignment_block:

```
 US-09-303-518D-649 x AAB01835   ..

 Align seg 1/1  to: AAB01835  from: 1  to: 1095

 574 CCTGACCGTGTTCGTATTGGGGCAGGCAGGCAATATTGGCGATCTGATGA 623
     ||||||:::|||   ||||||||||||    :::      |||||||||::
   1 ProAspGluValThrIleGlyAlaGly...AspValGlyArgSerAspAs 16

 624 AGAT............................GAGCCCAATAACCGCG 643
     ::::                              |||   |||:::    :
  16 pSerSerAspThrAlaPheProThrGlyThrGlyGluArgAsnSerProL 33

 644 AAAGTTCATATCATATTGCAAGTGCGTATTCTTGGCTCGTTGGTGGCAAT 693
     ::::::::   :::            :::   :::             |||
  33 ysThrAsnAlaGlnAsnArgProThrIleThr..............Asn 44

 694 ACCTTTGCACAAAATGGATCAGGTGGTGGCACA...GTCAACTTAGGTAG 740
     |||      ::::::          :::||||||    ||||||:::    ::
  45 ThrSerLeuGluGlnIleLeuLysAsnGlyThrPheValAsnIleThrAl 61

 741 TGAA..........................AAAATTAAACATAGCCCAT 763
```

```
        ::::                                 :::||||||::::::   :
  61 aLysAsnLysIleLeuValAsnSerAspIleAsnIleLysGluAsnSerH 78

 764 AT...............................GGTTTTTTACCAACA 780
     ::                                    |||
  78 isLeuIleLeuTrpSerGluArgAspGlyAsnSerGlyValGlnIleAsp 94

 781 GGAGGCTCATTTGGCGACAGTGGCTCACCAATGTTTATCTAT........ 822
     |||:::       :::   :::|||:::   :::   :::|||
  95 GlyAsnIleThrSerAlaThrGlyGlySerLeuThrValTyrSerSerGl 111

 823 .......GATGCCCAAAAGCAAAAGTGGTTAATTAATGGGGTATTG.... 861
            |||::::::|||:::         |||    :::|||   |||
 111 yTrpValAspValHisLysAsnIleThrLeuAsnSerGlyTyrLeuAsnI 128

 862 ........CAAACGGGCAACCCCTATATAGGAAAAAGCAATGGCTTCCAG 903
            ::::::|||:::        :::   ::::::|||    :::
 128 leThrThrLysSerGlyAspValAlaPheGluGlnGlyAsnAspLeuThr 144

 904 CTGGTT.......................CGTAAAGATTGGTTCTATGA 929
     ::::::                            :::|||   :::   :::||
 145 IleThrGlyGlnGlyThrIleThrAlaSerLysLysGlyPheArgPheAs 161

 930 TGAAATCTTTGCTGGAGATACCCATTCAGTATTC.........TACGAAC 970
     |::::::        :::   :::   :::   |||        |||:::
 161 pAsnValThrLeuSerGlyValLysLysGlyPheLeuPheLysTyrSerG 178

 971 CACGTCAAAATGGGAAATACTCT...............TTTAACGACGAT 1005
         :::|||:::|||   |||                     |||:::
 178 lnThrAsnAsnAsnLysAspSerAsnPheGluAsnHisPheArgGlyThr 194

1006 AATAATGGCACAGGAAAAATC............................ 1026
       |||    :::||||||:::
 195 LeuAsnIleSerGlyLysValAspIleLeuMetGlnAlaArgGlnGluAs 211

1027 ....AATGCCAAACATGAACACAATTCTCTGCCTAAT......AGATTAA 1066
         |||    :::|||:::   :::|||        |||      ||||||:
 211 nTrpAsnArgArgHisSerGlyArgSerHisTrpAsnValThrArgLeuA 228

1067 AAACACGAACCGTTCAATTGTTTAATGTTTCTTTATCCGAGACAGCAAGA 1116
     :::::    |||    :::    :::|||::::::::::::::::::::::
 228 snValSerThrAsnSerTyrLeuAsnIleThrIleAspAsnSerGlySer 244

1117 GAACCTGTTTATCATGCTGCAGGTGGTGTCAACAGTTATCGACCCAGACT 1166
     :::|||          :::   |||:::          ||||||
 245 ArgPro.........SerProGlyAlaGlyProLeuTyrArgArgSerGl 258

1167 GAATAATGGA.........GAAAATATTTCCTTTATTGACGAAGGAAAAG 1207
          ||||||          :::|||   :::   ::::::   ::::::|
 258 yLeuAsnGlyIleSerPheAsnAsnAspThrValPheAsnValAlaSerG 275

1208 GCGAATTGATA....................CTTACCAGCAACATCAAT 1236
     ||:::   :::                       ::::::||||||::::::
 275 lySerAlaValAsnPheSerIleLysProProIleValSerAsnValHis 291

1237 CAAGGTGCTGGAGGATTATATTTCCAAGGAGATTTTACGGTCTCGCCTGA 1286
     :::|||        |||   |||:::|||:::   :::|||
 292 AspGlyAsnHisThrLeu...PheAsnGlyAsnValSerVal........ 304

1287 AAATAACGAAACTTGGCAAGGCGCGGGC.........GTTCATATCAGTG 1327
                       |||:::|||          |||::::::
 305 ................LeuGlyGlyGlyAspValAsnPheHisPheAsnA 316

1328 AAGACAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAAACGACCGCCTG 1377
        :::|||:::      |||      :::   |||   :::::::::
 316 laSerSerSerAsnHisTrpThrHisGlyValValIleLysSerGlnAsn 332

1378 TCCAAAATCGGCAAAGGC...ACGCTGCACGTTCAAGCCAAAGGG..... 1419
        :::     ::::::|||   :::|||:::   :::::::::|||
 333 PheAsnAlaSerGluGlySerSerLeuArgPheLysSerGluGlySerTh 349

1420 ...................GAAAACCAAGGCTCGATCAGCGTGGGCGACG 1450
                        |||::::::   ::::::::::::      |
```

```
 349 rArgThrAlaPheThrIleGluSerAspLeuThrLeuAsnAlaThrGlyG 366

1451 GTACAGTCATTTTGGATCAGCAGGCAGACGATAAAGGCAAAAAACAAGCC 1500
     ||::::::   |||:::|||   |||          |||:::   |||
 366 lyAsnIleSerLeuAsnGlnValAlaGlyIleAspGlyAsnLeuGlnLys 382

1501 TTTAGTGAAATCGGCTTGGTCAGCGGCAGG.................GG 1532
                 :::||||||:::::::::                  ||
 383 ...........SerLeuValAlaAsnLysAsnIleThrPheGluGlyGl 395

1533 TACGGTGCAACTGAATGCCGATAATCAGTTCAACCCCGACAAACTCTATT 1582
     |::::::   |||   |||                  ||||||      :
 395 yAsnIleThrLeuAlaAla.................AspLysLysProI 406

1583 TCGGCTTTCGCGGCGGACGTTTGGATTTAAACGGGCATTCGCTTTCGTTC 1632
     ::    ::::::|||:::                 :::|||   ::::::::::::
 406 leGluIleLysGlyAsnIleThrValLysGluGlyAlaAsnValThrLeu 422

1633 CACCGTATTCAA...AATACCGATGAAGGGGCGATGATTGTCAACCACAA 1679
     :::      :::   ::::::|||::::::|||:::   ::::::   ||
 423 ArgSerAlaAsnTyrGlyAsnAspLysSerAlaLeuSerIleArgGlyAs 439

1680 TCAAGACAAAGAATCCACCGTTACCATTACAGGCAAT............ 1716
     |      :::::::::::::::|||:::||||||:::
 439 nValThrAsnLysGlyAsnLeuThrValThrGlySerAlaIleAsnIleG 456

1717 ..AAAGATATTGCTACAACCGGCAAT....................... 1740
       |||::::::::::::   |||:::
 456 luLysAsnLeuThrValGluGlySerAlaLysPheLeuAlaAsnProAsn 472

1741 .........................AACAACAGCTTGGATAGCAAAAA 1763
                               :::|||:::      |||:::
 473 TyrSerPheAsnValSerGlyLeuPheAspAsnGlnGlyLysSerAsnIl 489

1764 AGAAATTGCCTACAACGGT...TGGTTTGGCGAGAAAGATACGACCAAAA 1810
        :::||||||   :::|||      |||   :::   ::::::||||||:
 489 eSerIleAlaLysGlyGlyAlaHisPheLysAspIleAsnAsnThrLysS 506

1811 CGAACGGGCGGCTCAACCTTGTTTACCAGCCCGCCGCAGAAGACCGCACC 1860
     ::                |||||::::::   :::      :::   ||||||
 506 er.........LeuAsnIleThrThrAsnSerAspSerAlaTyrArgThr 519

1861 CTGCTGCTTTCCGGCGGAACAAATTTAAACGGCAACATCACGCAAACAAA 1910
     ::::::   ::::::   ||||||   ||||||:::::::::   |||:::
 520 IleIleGluGlyAsnIleThrAsnSerAsnGlyAspLeuAsnIleThrAs 536

1911 CGGCAAACTGTTTTTCAGCGGCAGACCAACACCGCACGCCTACAATCATT 1960
     ::::|||                          :::|||      ::::
 536 pAsnLys......................AsnAsnAlaGluIleGlnI 545

1961 TAAACGACCATTGGTCGCAAAAAGAGGGCATTCCTCGCGGGGAAATCGTG 2010
     :::::    :::   ||||||||||||                 |||:::::::::
 545 leGlyGlyAsnIleSerGlnLysGlu...........GlyAsnLeuThr 557

2011 TGGGACAACGACTGGATCAACCGCACATTTAAAGCGGAAAACTTCCAAAT 2060
       ::::::|||   ||||||   |||              :::::::::   ||
 558 IleSerSerAspLysIleAsnIleThr.........AsnGlnIleThrIl 571

2061 TAAAGGCGGA..................................... 2070
     ||||    |||
 571 eLysLysGlyValAsnLysGluAspSerAspSerSerThrAlaAsnAsnA 588

2071 ..CAGGCGGTGGTTTCCCGCAATCTTGCCAAAGTGAAAGGCGATTGGCAT 2118
        :::   :::::::::   :::      ::::::   ||||||   :::
 588 laAsnLeuThrIleLysThrLysGluLeuGlnLeuThrGlyAspLeuAsn 604

2119 TTGAGCAATCACGCCCAAGCAGTTTTTGGTGTCGCACCGCATCAAAGCCA 2168
     :::|||:::       :::|||   :::   :::|||   :::      :::
 605 IleSerGlyPheAspLysAlaGluIle...ThrAlaLysGluGlyAlaAs 620

2169 CACAATCTGTACACGTTCGGACTGGACGGGTCTGACAAATTGTGTCGAAA 2218
         |||      ::::||||||   ::::::   :::|||:::   :::
 620 pLeuIleIleGlyAsnSerAspAsnAsnAsnAsnAlaAsnAlaLysLysV 637
```

```
2219 AAACC.........ATTACCGACGATAAAGTGATTGCTTCATTGACTAAG 2259
       |||          :::   |||:::|||:::   |||:::      :::
 637 alThrPheAsnGlnValLysAspSerLysIleSerAlaAspSerHisAsn 653

2260 ACCGACATCAGCGGCAATGTCGATCTTGCC.................... 2289
     :::   ::::::::::::|||:::   :::
 654 ValThrLeuAsnSerLysValGluThrSerAsnGlyAsnAsnAspAlaGl 670

2290 .............GATCACGCTCATTTAAATCTCACAGGGCTT...GCCA 2323
                  |||   :::   |||::::::::::::         |
 670 uSerAsnAsnGlyAspGlyThrSerLeuThrIleAsnAlaLysAsnIleT 687

2324 CACTCAACGGCAATCTTAGTGCAAATGGCGATACACGTTATACAGTCAGC 2373
     ||:::|||:::|||:::::::::::::         ::::::   |||:::|||
 687 hrValAsnAsnAsnIleThrSerHisLysThrValAsnIleThrAlaSer 703

2374 CACAACGCCACCCAAAACGGCAACCTTAGCCTC.........GTGGGCAA 2414
     :::|||:::|||   :::::::::   ::::::         :::|||::
 704 GluAsnValThrThrLysAlaGlyThrThrIleAsnAlaThrThrGlySe 720

2415 TGCCCAAGCAACATTTAATCAAGCCACATTAAACGGCAACACA...TCGG 2461
     ::::::::::|||   :::   :::   ::::::|||::::::   |||:
 720 rValGluValThrAlaLysThrGlyAspIleLysGlyLysValGluSerT 737

2462 CTTCGGGCAATGCTTCATTTAATCTAAGCGACCACGCC............ 2499
     ::||||||::::::::::::::::   :::   :::|||
 737 hrSerGlySerValThrLeuThrAlaThrGlyGluAlaLeuAlaValSer 753

2500 ...GTACAAAACGGCAGTCTGACGCTTTCCGGCAAC.............. 2532
        :::::::::::::::::|||::::::::::|||
 754 AsnIleSerGlyAsnThrValThrIleThrAlaAsnLysGlyLysLeuTh 770

2533 .GCTAAGGCAAACGTAAGCCATTCCGCACTCAACGGTAATGTCTCCCTAG 2581
      ::::::|||:::   :::   ||||||:::||||||   ::::::   :
 770 rThrGlnAlaGlySerThrValSerAlaIleAsnGlyValThrAlaSerS 787

2582 CCGATAAGGCAGTATTCCATTTTGAAAGCAGCCGCTTTACCGGACAAATC 2631
     ::                          :::|||:::   ::::::|||   |||
 787 er.....................GlnSerGlyAspIleSerGlyThrIle 796

2632 AGCGGCGGC......AAGGATACGGCATTACACTTAAAAGACAGCGAATG 2675
     ||||||:::         |||   :::|||:::          ::::::
 797 SerGlyAsnThrValLysValSerAlaIle............GlyAspLe 809

2676 GACGCTGCCGTCAGGCACGGAATTA............GGCAATTTAAACC 2713
       |||      ||||||:::|||:::              |||:::   |||:
 809 uThrThrLysSerGlySerGluIleLysAlaLysThrGlyGluAlaAsnV 826

2714 TTGACAACGCCACCATTACACTCAATTCCGCCTATCGCCACGATGCGGCA 2763
     ::   :::||||||   |||::::::::::::         :::|||:::
 826 alThrSerAlaThrGlyThrIleGlyGlyThrIleSerGlyAsnAlaVal 842

2764 GGG......GCGCAAACCGGCAGTGCGACA......GATGCGCCGCGCCG 2801
     :::         |||:::||||||:::   |||      ||||||   :::
 843 AsnValThrAlaAsnThrGlyAspLeuThrValGluAspAlaAlaLysIl 859

2802 CCGTTCGCGCCGTTCGCGCCGTTCCCTATTATCCGTTACACCGCCAACTT 2851
        :::      :::      :::|||   :::::::::           :
 859 eAspAlaThrGlyGlyAlaAlaThrLeuThrAlaThrSerGlyLysLeuT 876

2852 CGGTAGAATCCCGTTTCAACACGCTGACGGTAAACGGCAAATTGAACGGT 2901
     ::::::::::         :::      ::::::|||:::::::::|||
 876 hrThrLysAlaSerSerSerIleThrSerAlaAsnAsnGlnValAsn... 891

2902 CAGGGAACATTCCGCTTTATGTCGGAACTCTTCGGCTACCGCAGCGACAA 2951

 891 .................................................. 891

2952 ATTGAAGCTGGCGGAAAGTTCCGAAGGCACTTACACCTTGGCGGTCAACA 3001
       |||:::         ::::::|||   |||              |||:::|||
 892 .LeuSerAlaLysAspGlySerIleGlyGlyAsnIleAsnAlaAlaAsnV 908

3002 ATACCGGCAACGAACCTGCAAGCCTCGAACAATTG............... 3036
       |||   |||      :::::::|||         :::
 908 alThrLeuAsnThrThrGlyAlaLeuThrThrValLysGlySerSerIle 924

3037 ..................ACGGTAGTGGAAGGAAAAGAC.........AA 3059
                       |||:::::::::||||||            ||
 925 AsnAlaAsnSerGlyThrLeuValIleAsnAlaLysAspAlaGluLeuAs 941

3060 CAAACCGCTGTCCGAAAAC.........CTTAATTTCACCCTGCAAAACG 3100
     |           |||   |||         :::|||   |||      |||
 941 nGlyGluAlaSerGlyAsnHisThrValValAsnAlaThrAsnAlaAsnG 958

3101 AACACGTCGATGCCGGCGCGTGGCGTTACCAACTCATCCGCAAAGACGGC 3150
             ::::::   |||         :::   ::::::      |||
 958 lySerGlySerValIleAlaThrThrSerSerArgValAsnIleThrGly 974

3151 GAGTTCCGCCTGCATAATCCGGTCAAAGAACAAGAGCTTTCCGACAAACT 3200
     ::::::         |||   :::         ::::::   :::|||
 975 AspLeuIleThrIleAsnGlyLeu.........AsnIleIleSerLysAs 988

3201 CGGCAAGGCAGAAGCCAAAAAACAGGCGGAAAAAGACAACGCGCAA.... 3246
       |||         :::      ::::::   |||   :::::::::
 988 nGlyIleAsnThrValLeuLeuLysGlyValLysIleAspValLysTyrI 1005

3247 .................AGCCTTGACGCGCTGATTGCGGCCGGGCGCGAT 3279
                       |||:::|||   :::|||   |||   |||
1005 leGlnProGlyIleAlaSerValAspGluValIleGluAlaLysArg... 1020

3280 GCCGTCGAAAAGACAGAAAGCGTTGCCGAACCGGCCCGGCAGGCAGGCGG 3329
        :::||||||:::::::::::::::::         |||:::|||   ::
1021 IleLeuGluLysValLysAspLeuSerAspGluGluArgGluAlaLeuAl 1037

3330 GGAA...AATGTCGGCATTATGCAGGCGGAGGAAGAGAAAAAACGGGTGC 3376
      ::::   :::|||:::   ::::::      |||   ::::::   :::
1037 aLysLeuGlyValSerAlaValArgPheAlaGluProAsnAsnAlaIle. 1053

3377 AGGCGGATAAAGACACCGCCTTGGCGAAACAGCGCGAAGCGGAAACCCGG 3426
                       ::::::   |||:::|||      ||||||
1054 .................ThrIleAsnThrGlnAsnGluPheThrThrArg 1064

3427 CCGGCTACC 3435
      |||::::::
1065 ProSerSer 1067
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01834

seq_documentation_block:

```
ID   AAB01834 standard; Protein; 1101 AA.
XX
AC   AAB01834;
XX
DT   11-SEP-2000  (first entry)
XX
DE   Haemophilus influenzae strain LCDC2 HMW1A protein, SEQ ID NO:43.
XX
KW   HMW protein; hmw gene; hmwA1; hmwA2; high molecular weight;
KW   non-typeable Haemophilus influenzae; NTHi; non-encapsulated;
KW   recombinant production; Escherichia coli; antibacterial; vaccine;
KW   human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;
KW   detection; diagnosis.
XX
OS   Haemophilus influenzae strain LCDC2.
XX
PN   WO200020609-A2.
XX
PD   13-APR-2000.
XX
PF   07-OCT-1999;   99WO-CA00938.
XX
PR   07-OCT-1998;   98US-0167568.
PR   08-DEC-1998;   98US-0206942.
XX
PA   (CONN-) CONNAUGHT LAB LTD.
```

```
XX
PI   Loosmore SM,  Yang Y,  Klein MH;
XX
DR   WPI; 2000-303789/26.
DR   N-PSDB; AAA52183.
XX
PT   Nucleic acid molecule for producing recombinant high molecular weight
PT   proteins of Haemophilus which are used as a vaccine to provide
PT   protection against Haemophilus induced diseases in humans -
XX
PS   Claim 12; Fig 22A-P; 307pp; English.
XX
CC   The invention relates to the recombinant production of Haemophilus
CC   influenzae high molecular weight (HMW) proteins in Escherichia coli. The
CC   expression construct used to effect recombinant expression comprises a
CC   promoter functional in E. coli (e.g., the T7 promoter) operably linked
CC   to a modified hmwABC operon from a non-typeable (non-encapsulated) H.
CC   influenzae (NTHi). Most HMW-expressing NTHi strains contain two hmw gene
CC   clusters termed hmw1ABC and hmw2ABC. Each hmwABC operon comprises hmwA,
CC   hmwB and hmwC genes. The hmwA genes encode the structural HMWA proteins
CC   and the hmwB and hmwC genes encode accessory proteins which are
CC   responsible for post-translational processing and secretion of the HMWA
CC   proteins. The modified hmwABC operon used in the expression construct of
CC   the invention contains an A gene modified such that it encodes only the
CC   mature HMWA. The invention also discloses hmwA genes (AAA52175-A52198)
CC   and HMWA proteins (AAB01824-B01849) from the non-typeable H. influenzae
CC   strains Joyc, K1, K21, LCDC2, PMH1, 15 and 12. The nucleic acids and
CC   vectors are used for the production of recombinant H. influenzae HMW
CC   proteins which can be used as vaccines to mediate a humoral or
CC   cell-mediated immune response to provide protection against diseases in
CC   humans caused by H. influenzae (e.g., otitis media, epiglottitis,
CC   pneumonia and tracheobronchitis). The HMW proteins are also useful as
CC   antigens in immunoassays for detecting antibodies against Haemophilus,
CC   HMW proteins and/or HMW peptides. The nucleotide sequences encoding the
CC   HMW proteins can be used to isolate and clone hmw genes from other
CC   non-typeable strains of Haemophilus via hybridisation reactions. The
CC   present sequence represents an HMWA protein from a non-typeable strain of
CC   H. influenzae.
XX
SQ   Sequence   1101 AA;

alignment_scores:
          Quality:  264.50         Length:   1153
            Ratio:  0.442            Gaps:     56
 Percent Similarity:  51.865  Percent Identity: 20.035

alignment_block:
 US-09-303-518D-649 x AAB01834   ..

 Align seg 1/1  to: AAB01834  from: 1  to: 1101

    574 CCTGACCGTGTTCGTATTGGGGCAGGCAGGCAATATTGGCGATCTGATGA 623
        ||||||:::|||   |||||||||||    :::      |||||||||::
      7 ProAspGluValThrIleGlyAlaGly...AspValGlyArgSerAspAs 22

    624 AGAT............................GAGCCCAATAACCGCG 643
        ::::                            |||  |||:::   :
     22 pSerSerAspThrAlaPheProThrGlyThrGlyGluArgAsnSerProL 39

    644 AAAGTTCATATCATATTGCAAGTGCGTATTCTTGGCTCGTTGGTGGCAAT 693
        :::::::   :::             :::   :::            |||
     39 ysThrAsnAlaGlnAsnArgProThrIleThr..............Asn 50

    694 ACCTTTGCACAAAATGGATCAGGTGGTGGCACA...GTCAACTTAGGTAG 740
        |||       ::::::          :::||||||  |||||||:::   ::
     51 ThrSerLeuGluGlnIleLeuLysAsnGlyThrPheValAsnIleThrAl 67

    741 TGAA..........................AAAATTAAACATAGCCCAT 763
        ::::                          :::||||||:::::    :
     67 aLysAsnLysIleLeuValAsnSerAspIleAsnIleLysGluAsnSerH 84

    764 AT..............................GGTTTTTTACCAACA 780
        ::                                             |||
     84 isLeuIleLeuTrpSerGluArgAspGlyAsnSerGlyValGlnIleAsp 100

    781 GGAGGCTCATTTGGCGACAGTGGCTCACCAATGTTTATCTAT........ 822
        |||:::        :::   :::|||:::   :::    :::|||
    101 GlyAsnIleThrSerAlaThrGlyGlySerLeuThrValTyrSerSerGl 117

    823 .......GATGCCCAAAAGCAAAAGTGGTTAATTAATGGGGTATTG.... 861
               |||::::::|||:::        |||   :::|||   |||
    117 yTrpValAspValHisLysAsnIleThrLeuAsnSerGlyTyrLeuAsnI 134

    862 ........CAAACGGGCAACCCCTATATAGGAAAAAGCAATGGCTTCCAG 903
                ::::::|||:::      :::   :::::::|||    :::
    134 leThrThrLysSerGlyAspValAlaPheGluGlnGlyAsnAspLeuThr 150

    904 CTGGTT......................CGTAAAGATTGGTTCTATGA 929
        ::::::                      :::|||   :::    :::||
    151 IleThrGlyGlnGlyThrIleThrAlaSerLysLysGlyPheArgPheAs 167

    930 TGAAATCTTTGCTGGAGATACCCATTCAGTATTC.........TACGAAC 970
        |::::::      :::   :::   :::   |||         |||:::
    167 pAsnValThrLeuSerGlyValLysLysGlyPheLeuPheLysTyrSerG 184

    971 CACGTCAAAATGGGAAATACTCT...............TTTAACGACGAT 1005
           :::|||:::|||     |||                |||:::
    184 lnThrAsnAsnAsnLysAspSerAsnPheGluAsnHisPheArgGlyThr 200

   1006 AATAATGGCACAGGAAAAATC............................ 1026
          |||    :::||||||:::
    201 LeuAsnIleSerGlyLysValAspIleLeuMetGlnAlaArgGlnGluAs 217

   1027 ....AATGCCAAACATGAACACAATTCTCTGCCTAAT......AGATTAA 1066
            |||    :::|||:::    :::|||       |||      ||||||:
    217 nTrpAsnArgArgHisSerGlyArgSerHisTrpAsnValThrArgLeuA 234

   1067 AAACACGAACCGTTCAATTGTTTAATGTTTCTTTATCCGAGACAGCAAGA 1116
        :::::   |||    :::    :::|||::::::::::::::::::::::
    234 snValSerThrAsnSerTyrLeuAsnIleThrIleAspAsnSerGlySer 250

   1117 GAACCTGTTTATCATGCTGCAGGTGGTGTCAACAGTTATCGACCCAGACT 1166
        :::|||            :::    |||:::           ||||||
    251 ArgPro.........SerProGlyAlaGlyProLeuTyrArgArgSerGl 264

   1167 GAATAATGGA.........GAAAATATTTCCTTTATTGACGAAGGAAAAG 1207
          ||||||          :::|||    :::    ::::::   ::::::|
    264 yLeuAsnGlyIleSerPheAsnAsnAspThrValPheAsnValAlaSerG 281

   1208 GCGAATTGATA....................CTTACCAGCAACATCAAT 1236
        ||:::    :::                   ::::::||||||::::::
    281 lySerAlaValAsnPheSerIleLysProProIleValSerAsnValHis 297

   1237 CAAGGTGCTGGAGGATTATATTTCCAAGGAGATTTTACGGTCTCGCCTGA 1286
        :::|||          |||   |||:::|||:::    :::|||
    298 AspGlyAsnHisThrLeu...PheAsnGlyAsnValSerVal........ 310

   1287 AAATAACGAAACTTGGCAAGGCGCGGGC.........GTTCATATCAGTG 1327
                             |||:::|||           |||::::::
    311 ................LeuGlyGlyGlyAspValAsnPheHisPheAsnA 322

   1328 AAGACAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAAACGACCGCCTG 1377
          :::|||:::       |||        :::   |||   :::::::::
    322 laSerSerSerAsnHisTrpThrHisGlyValValIleLysSerGlnAsn 338

   1378 TCCAAAATCGGCAAAGGC...ACGCTGCACGTTCAAGCCAAAGGG..... 1419
          :::   ::::::|||    :::|||:::    ::::::::::|||
    339 PheAsnAlaSerGluGlySerSerLeuArgPheLysSerGluGlySerTh 355

   1420 ..................GAAAACCAAGGCTCGATCAGCGTGGGCGACG 1450
                          |||::::::    ::::::::::::     |
    355 rArgThrAlaPheThrIleGluSerAspLeuThrLeuAsnAlaThrGlyG 372

   1451 GTACAGTCATTTTGGATCAGCAGGCAGACGATAAAGGCAAAAAACAAGCC 1500
        ||::::::   |||:::|||    |||           |||:::   |||
```

```
 372 lyAsnIleSerLeuAsnGlnValAlaGlyIleAspGlyAsnLeuGlnLys 388

1501 TTTAGTGAAATCGGCTTGGTCAGCGGCAGG..................GG 1532
                 :::||||||:::::::::                   ||
 389 ............SerLeuValAlaAsnLysAsnIleThrPheGluGlyGl 401

1533 TACGGTGCAACTGAATGCCGATAATCAGTTCAACCCCGACAAACTCTATT 1582
     |:::::   |||   |||                   ||||||      :
 401 yAsnIleThrLeuAlaAla.................AspLysLysProI 412

1583 TCGGCTTTCGCGGCGGACGTTTGGATTTAAACGGGCATTCGCTTTCGTTC 1632
     ::   ::::::|||:::               :::|||   :::::::::::
 412 leGluIleLysGlyAsnIleThrValLysGluGlyAlaAsnValThrLeu 428

1633 CACCGTATTCAA...AATACCGATGAAGGGGCGATGATTGTCAACCACAA 1679
     :::       :::   ::::::|||::::::|||:::   ::::::   ||
 429 ArgSerAlaAsnTyrGlyAsnAspLysSerAlaLeuSerIleArgGlyAs 445

1680 TCAAGACAAAGAATCCACCGTTACCATTACAGGCAAT............. 1716
     |      ::::::::::::::|||:::|||||||:::
 445 nValThrAsnLysGlyAsnLeuThrValThrGlySerAlaIleAsnIleG 462

1717 ..AAAGATATTGCTACAACCGGCAAT........................ 1740
        |||:::::::::::    |||:::
 462 luLysAsnLeuThrValGluGlySerAlaLysPheLeuAlaAsnProAsn 478

1741 ..........................AACAACAGCTTGGATAGCAAAAA 1763
                               :::|||:::     |||:::
 479 TyrSerPheAsnValSerGlyLeuPheAspAsnGlnGlyLysSerAsnIl 495

1764 AGAAATTGCCTACAACGGT...TGGTTTGGCGAGAAAGATACGACCAAAA 1810
      :::||||||    :::|||       |||   :::   ::::::||||||:
 495 eSerIleAlaLysGlyGlyAlaHisPheLysAspIleAsnAsnThrLysS 512

1811 CGAACGGGCGGCTCAACCTTGTTTACCAGCCCGCCGCAGAAGACCGCACC 1860
     ::            ||||||::::::   :::      :::     ||||||
 512 er.........LeuAsnIleThrThrAsnSerAspSerAlaTyrArgThr 525

1861 CTGCTGCTTTCCGGCGGAACAAATTTAAACGGCAACATCACGCAAACAAA 1910
     ::::::   ::::::   ||||||   ||||||:::::::::   |||::
 526 IleIleGluGlyAsnIleThrAsnSerAsnGlyAspLeuAsnIleThrAs 542

1911 CGGCAAACTGTTTTTCAGCGGCAGACCAACACCGCACGCCTACAATCATT 1960
     ::::|||                          :::|||      :::
 542 pAsnLys.......................AsnAsnAlaGluIleGlnI 551

1961 TAAACGACCATTGGTCGCAAAAAGAGGGCATTCCTCGCGGGGAAATCGTG 2010
     :::::   :::   ||||||||||||          |||:::::::::::
 551 leGlyGlyAsnIleSerGlnLysGlu............GlyAsnLeuThr 563

2011 TGGGACAACGACTGGATCAACCGCACATTTAAAGCGGAAAACTTCCAAAT 2060
        ::::::|||   ||||||   |||          :::::::::   ||
 564 IleSerSerAspLysIleAsnIleThr.........AsnGlnIleThrIl 577

2061 TAAAGGCGGA........................................ 2070
     ||||   |||
 577 eLysLysGlyValAsnLysGluAspSerAspSerSerThrAlaAsnAsnA 594

2071 ..CAGGCGGTGGTTTCCCGCAATGTTGCCAAAGTGAAAGGCGATTGGCAT 2118
       :::   :::::::::   :::      ::::::   ||||||   :::
 594 laAsnLeuThrIleLysThrLysGluLeuGlnLeuThrGlyAspLeuAsn 610

2119 TTGAGCAATCACGCCCAAGCAGTTTTTGGTGTCGCACCGCATCAAAGCCA 2168
     :::|||:::       :::|||   :::    :::|||   :::   :::
 611 IleSerGlyPheAspLysAlaGluIle...ThrAlaLysGluGlyAlaAs 626

2169 CACAATCTGTACACGTTCGGACTGGACGGGTCTGACAAATTGTGTCGAAA 2218
        |||      :::||||||   ::::::   :::|||:::   :::
 626 pLeuIleIleGlyAsnSerAspAsnAsnAsnAsnAlaAsnAlaLysLysV 643

2219 AAACC.........ATTACCGACGATAAAGTGATTGCTTCATTGACTAAG 2259
        |||          :::   |||:::|||:::   |||:::      :::
 643 alThrPheAsnGlnValLysAspSerLysIleSerAlaAspSerHisAsn 659

2260 ACCGACATCAGCGGCAATGTCGATCTTGCC.................... 2289
     :::   ::::::::::::|||:::   :::
 660 ValThrLeuAsnSerLysValGluThrSerAsnGlyAsnAsnAspAlaGl 676

2290 .............GATCACGCTCATTTAAATCTCACAGGGCTT...GCCA 2323
                  |||   :::   |||:::::::::::::        |
 676 uSerAsnAsnGlyAspGlyThrSerLeuThrIleAsnAlaLysAsnIleT 693

2324 CACTCAACGGCAATCTTAGTGCAAATGGCGATACACGTTATACAGTCAGC 2373
     ||:::|||:::|||:::::::::::         ::::::   |||:::|||
 693 hrValAsnAsnAsnIleThrSerHisLysThrValAsnIleThrAlaSer 709

2374 CACAACGCCACCCAAAACGGCAACCTTAGCCTC.........GTGGGCAA 2414
     :::|||:::|||   :::::::::   ::::::         :::|||::
 710 GluAsnValThrThrLysAlaGlyThrThrIleAsnAlaThrThrGlySe 726

2415 TGCCCAAGCAACATTTAATCAAGCCACATTAAACGGCAACACA...TCGG 2461
     :::::::::|||   :::   :::   ::::::|||::::::   |||:
 726 rValGluValThrAlaLysThrGlyAspIleLysGlyLysValGluSerT 743

2462 CTTCGGGCAATGCTTCATTTAATCTAAGCGACCACGCC............ 2499
     :::||||||:::::::::::::::   :::   :::|||
 743 hrSerGlySerValThrLeuThrAlaThrGlyGluAlaLeuAlaValSer 759

2500 ...GTACAAAACGGCAGTCTGACGCTTTCCGGCAAC.............. 2532
        ::::::::::::::::|||:::::::::|||
 760 AsnIleSerGlyAsnThrValThrIleThrAlaAsnLysGlyLysLeuTh 776

2533 .GCTAAGGCAAACGTAAGCCATTCCGCACTCAACGGTAATGTCTCCCTAG 2581
       ::::::|||:::   :::   ||||||:::||||||   ::::::   :
 776 rThrGlnAlaGlySerThrValSerAlaIleAsnGlyValThrAlaSerS 793

2582 CCGATAAGGCAGTATTCCATTTTGAAAGCAGCCGCTTTACCGGACAAATC 2631
     ::                            :::|||:::   ::::::|||   |||
 793 er....................GlnSerGlyAspIleSerGlyThrIle 802

2632 AGCGGCGGC......AAGGATACGGCATTACACTTAAAAGACAGCGAATG 2675
     ||||||:::       |||   :::|||:::           ::::::
 803 SerGlyAsnThrValLysValSerAlaIle............GlyAspLe 815

2676 GACGCTGCCGTCAGGCACGGAATTA...........GGCAATTTAAACC 2713
      |||      ||||||:::|||:::              |||:::   |||:
 815 uThrThrLysSerGlySerGluIleLysAlaLysThrGlyGluAlaAsnV 832

2714 TTGACAACGCCACCATTACACTCAATTCCGCCTATCGCCACGATGCGGCA 2763
     ::   :::||||||   |||::::::::::::         :::|||:::
 832 alThrSerAlaThrGlyThrIleGlyGlyThrIleSerGlyAsnAlaVal 848

2764 GGG......GCGCAAACCGGCAGTGCGACA......GATGCGCCGCGCCG 2801
     :::         |||:::|||||||:::   |||      ||||||   :::
 849 AsnValThrAlaAsnThrGlyAspLeuThrValGluAspAlaAlaLysIl 865

2802 CCGTTCGCGCCGTTCGCGCCGTTCCCTATTATCCGTTACACCGCCAACTT 2851
        :::      :::      :::|||   :::::::::            :
 865 eAspAlaThrGlyGlyAlaAlaThrLeuThrAlaThrSerGlyLysLeuT 882

2852 CGGTAGAATCCCGTTTCAACACGCTGACGGTAAACGGCAAATTGAACGGT 2901
     ::::::::::::      :::      ::::::|||::::::::::|||
 882 hrThrLysAlaSerSerSerIleThrSerAlaAsnAsnGlnValAsn... 897

2902 CAGGGAACATTCCGCTTTATGTCGGAACTCTTCGGCTACCGCAGCGACAA 2951

 897 .................................................. 897

2952 ATTGAAGCTGGCGGAAAGTTCCGAAGGCACTTACACCTTGGCGGTCAACA 3001
      |||:::         :::::|||   |||            |||:::|||
 898 .LeuSerAlaLysAspGlySerIleGlyGlyAsnIleAsnAlaAlaAsnV 914

3002 ATACCGGCAACGAACCTGCAAGCCTCGAACAATTG............... 3036
       |||   |||      :::::::|||         :::
 914 alThrLeuAsnThrThrGlyAlaLeuThrThrValLysGlySerSerIle 930
```

```
3037 ..................ACGGTAGTGGAAGGAAAAGAC.........AA 3059
                       |||:::::::::||||||         ||
 931 AsnAlaAsnSerGlyThrLeuValIleAsnAlaLysAspAlaGluLeuAs 947

3060 CAAACCGCTGTCCGAAAAC.........CTTAATTTCACCCTGCAAAACG 3100
     |          |||    |||       :::|||   |||      |||
 947 nGlyGluAlaSerGlyAsnHisThrValValAsnAlaThrAsnAlaAsnG 964

3101 AACACGTCGATGCCGGCGCGTGGCGTTACCAACTCATCCGCAAAGACGGC 3150
              ::::::   |||          :::   ::::::      |||
 964 lySerGlySerValIleAlaThrThrSerSerArgValAsnIleThrGly 980

3151 GAGTTCCGCCTGCATAATCCGGTCAAAGAACAAGAGCTTTCCGACAAACT 3200
     ::::::          |||   :::          ::::::   :::|||
 981 AspLeuIleThrIleAsnGlyLeu.........AsnIleIleSerLysAs 994

3201 CGGCAAGGCAGAAGCCAAAAAACAGGCGGAAAAAGACAACGCGCAA.... 3246
       |||        :::     ::::::   |||   :::::::::
 994 nGlyIleAsnThrValLeuLeuLysGlyValLysIleAspValLysTyrI 1011

3247 .................AGCCTTGACGCGCTGATTGCGGCCGGGCGCGAT 3279
                      |||:::|||   :::|||   |||   |||
1011 leGlnProGlyIleAlaSerValAspGluValIleGluAlaLysArg... 1026

3280 GCCGTCGAAAAGACAGAAAGCGTTGCCGAACCGGCCCGGCAGGCAGGCGG 3329
        :::||||||:::::::::::::::::::      |||:::|||    ::
1027 IleLeuGluLysValLysAspLeuSerAspGluGluArgGluAlaLeuAl 1043

3330 GGAA...AATGTCGGCATTATGCAGGCGGAGGAAGAGAAAAAACGGGTGC 3376
     ::::   :::|||:::   ::::::       |||   ::::::   :::
1043 aLysLeuGlyValSerAlaValArgPheAlaGluProAsnAsnAlaIle. 1059

3377 AGGCGGATAAAGACACCGCCTTGGCGAAACAGCGCGAAGCGGAAACCCGG 3426
                      ::::::      |||:::|||       ||||||
1060 .................ThrIleAsnThrGlnAsnGluPheThrThrArg 1070

3427 CCGGCTACC 3435
     |||::::::
1071 ProSerSer 1073
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAU34143

seq_documentation_block:

```
ID   AAU34143 standard; Protein; 2086 AA.
XX
AC   AAU34143;
XX
DT   14-FEB-2002  (first entry)
XX
DE   Staphylococcus aureus cellular proliferation protein #419.
XX
KW   Antisense; prokaryotic cellular proliferation protein;
KW   antibiotic; antibacterial; drug design.
XX
OS   Staphylococcus aureus.
XX
PN   WO200170955-A2.
XX
PD   27-SEP-2001.
XX
PF   21-MAR-2001; 2001WO-US09180.
XX
PR   21-MAR-2000; 2000US-191078P.
PR   23-MAY-2000; 2000US-206848P.
PR   26-MAY-2000; 2000US-207727P.
PR   23-OCT-2000; 2000US-242578P.
PR   27-NOV-2000; 2000US-253625P.
PR   22-DEC-2000; 2000US-257931P.
PR   16-FEB-2001; 2001US-269308P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Haselbeck R,  Ohlsen KL,  Zyskind JW,  Wall D,  Trawick JD,  Carr GJ;
PI   Yamamoto RT,  Xu HH;
XX
DR   WPI; 2001-611495/70.
DR   N-PSDB; AAS52002.
XX
PT   New polynucleotides for the identification and development of
PT   antibiotics, comprise sequences of antisense nucleic acids -
XX
PS   Example 3; Seq ID No 5639; 511pp; English.
XX
CC   The invention relates to antisense inhibitors of genes essential to
CC   prokaryotic cellular proliferation, their use in identifying the
CC   genes, their use in the discovery of novel antibiotics, the essential
CC   genes themselves and the encoded proteins. The prokaryotes used are
CC   Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC   pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC   invention is also useful for the identification of potential new targets
CC   for antibiotic development. The antisense nucleic acids can also be used
CC   to identify proteins used in proliferation, to express these proteins,
CC   and to obtain antibodies capable of binding to the expressed proteins.
CC   The proteins can be used to screen compounds in rational drug discovery
CC   programmes. The antisense nucleic acid sequence is also useful to screen
CC   for homologous nucleic acids which are required for cell proliferation in
CC   a wide variety of organisms. The present sequence represents an
CC   essential prokaryotic cellular proliferation protein.
CC   Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic
CC   format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   2086 AA;
```

alignment_scores:

```
           Quality:  264.50            Length:   1388
             Ratio:  0.387               Gaps:     69
Percent Similarity:  49.207   Percent Identity: 20.605
```

alignment_block:

```
US-09-303-518D-649 x AAU34143   ..

Align seg 1/1  to: AAU34143  from: 1  to: 2086

 505 ACAGATGCAGAACCTGTTGAAATGACCAGTTAT...........ATGGA 542
     |||||||||:::|||   ::::::::::::|||             ::
 697 ThrAspAlaSerProAsnAsnGlnSerThrTyrAsnSerAlaValSerAs 713

 543 TGGGCGGAAATATATCGATCAAAATAATTACCCTGACCGTGTTCGTATTG 592
     :::::::      |||:::|||:::|||   |||
 713 nAlaLysGlyIleIleAsnGlnThrAsnAsnPro................ 724

 593 GGGCAGGCAGGCAATATTGGCGATCTGATGAAGATGAGCCCAATAACCGC 642

 724 .................................................. 724

 643 GAAAGTTCATATCATATTGCAAGTGCGTATTCTTGGCTCGTTGGTGGCAA 692
                         :::      |||   ::::::   :::::
 725 ..................ThrMetAspThrSerAlaIleThrGlnAlaTh 735

 693 TACCTTTGCACAAAATGGATCAGGTGGTGGCACAGTCAACTTAGGTAGTG 742
     :|||   ::::::|||:::::::::|||        :::|||   |||::|
 735 rThrGlnValAsnAsnAlaLysAsnGly......LeuAsn...GlyAlaG 749

 743 AAAAAATTAAACATAGCCCATATGGTTTTTTACCAACAGGAGGCTCATTT 792
     ||::::::::::: |||:::   :::       |||          :::||
 749 luAsnLeuArgAsn.AlaGlnAsnThrAlaLysGln.........AsnLe 762

 793 GGCGACAGTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCAAAAGTG 842
     |   |||:::::|||                :::   ::::::::::::
 762 uAsnThrLeuSerHis...............LeuThrAsnAsnGlnLysS 774

 843 GTTAATTAATGGGGTATTGCAAACGGGCAACCCCTATATAGGAAAAAGCA 892
     ::   :::   :::          ||||||      :::   |||   :::
```

```
 774 erAlaIleSerSerGlnIleAspArgAlaGlyHisValSerGluValThr 790

 893 ATGGCTTCCAGCTGGTTCGTAAAGATTGGTTCTATGATGAAATCTTTGCT 942
        |||:::::     :::::::::::::::::|||   :::|||
 791 AlaAlaLysAsnAlaAlaThrGluLeuAsnThrGlnMetGlyAsnLeuGl 807

 943 GGAGATACCCATTCAGTATTCTACGAACCACGTCAAAATGGGAAATACTC 992
      :::          ::::::  |||   ||||||:::|||     ::
 807 uGln......AlaIleHisAspGlnAsnThrValLysGlnGly..ValAs 821

 993 TTTTAACGACGATAATAATGGCACAGGAAAAATCAATGCCAAACATGAAC 1042
     :|||:::|||  :::                          |||  ::::
 821 nPheThrAspAlaAsp......................LysAlaLysA 830

1043 ACAATTCTCTGCCTAATAGATTAAAAACACGAACCGTTCAATTGTTTAAT 1092
     ::::::::      |||  ::::::          :::  |||   :::
 830 rgAspAlaTyrThrAsnAlaValSerArgAlaGluThrIleLeuAsnLys 846

1093 GTTTCTTTATCCGAGACAGCAAGAGAACCTGTTTATCATGCTGCAGGTGG 1142
     ::::::  :::::::|||:::::::::  |||       |||       ::
 847 ThrGlnGlyAlaAsnThrSerLysGlnAspValGluAlaAlaIleGlnAs 863

1143 TGTCAACAGTTATCGACCCAGACTGAATAATGGAGAAAATATTTCCTTTA 1192
     :|||:::|||   :::      |||||||:::   :::|||::::::
 863 nValThrSerAlaLysAsnAlaLeuAsnGlyAspGlnAsnValThrAsnA 880

1193 TT...GACGAAGGAAAAGGCGAATTGATACTTACCAGCAACATCAATCAA 1239
          :::   :::|||        |||       ::::::|||||||:::
 880 laLysAsnThrAlaLysHisAlaLeuAsnAsnLeuThrSerIleAsnAsn 896

1240 GGTGCTGGAGGATTATATTTCCAAGGAGATTTTACGGTCTCGCCTGAAAA 1289
     :::                    :::  |||:::|||::::::   :::::
 897 AlaGln..............LysArgAspLeuThrThrLysIleAspGl 908

1290 TAACGAAACTTGGCAAGGCGCGGGCGTTCATATCAGTGAAGACAGTACCG 1339
     :       |||     |||
 908 nAlaThrThrValAlaGly............................. 914

1340 TTACTTGGAAAGTAAACGGCGTGGCAAACGACCGCCTGTCCAAAATCGGC 1389
                   |||::::::|||:::|||
 915 ..........ValGluAlaValSerAsn.................... 920

1390 AAAGGCACGCTGCACGTTCAAGCCAAAGGGGAAAACCAAGGCTCGATCAG 1439
        ||||||             |||    ::::::   |||::::::|||::
 921 ThrGlyThrGlnLeuAsnThrAlaMetAlaAsnLeuGlnAsnGlyIleAs 937

1440 CGTGGGCGACGGTACAGTCATTTTGGATCAGCAGGCAGAC...GATAAAG 1486
     :           :::|||:::       ::::::        |||   |||:::
 937 nAspLysAlaAsnThrLeuAlaSerGluAsnTyrHisAspAlaAspSerA 954

1487 GCAAAAAACAAGCCTTTAGTGAAATCGGCTTGGTCAGCGGCAGGGGTACG 1536
        ||||||    |||:::::::::
 954 spLysLysThrAlaTyrThrGln.......................... 961

1537 GTGCAACTGAATGCCGATAATCAGTTCAACCCCGACAAACTCTATTTCGG 1586
     :::         |||||:::|||   :::|||   ::::::
 962 AlaValThrAsnAlaGluAsnIleLeuAsnLysAsnSer........... 974

1587 CTTTCGCGGCGGACGTTTGGATTTAAACGGGCATTCGCTTTCGTTCCACC 1636
              |||::::::||||||  :::  :::  ::::::
 975 .......GlySerAsnLeuAspLysAlaAlaValGluAsnAlaLeuSerG 989

1637 GTATTCAAAATACCGATGAAGGGGCGATGATTGTCAACCACAATCAAGAC 1686
     :::::  |||:::   :::||||||:::        |||||||||   :::
 989 lnValThrAsnAla...LysGlyAlaLeuAsnGlyAsnHisAsnLeuGlu 1004

1687 AAA.........GAATCCACCGTTACCATTACAGGCAATAAAGATATTGC 1727
     :::              ::::::::::::||||||:::|||   :::   :::::
1005 GlnAlaLysSerAsnAlaAsnThrThrIleAsnGlyLeuGlnHisLeuTh 1021

1728 TACA..................................ACCGGCAATA 1741
     :|||                                    :::   |||
1021 rThrAlaGlnLysAspLysLeuLysGlnGlnValGlnGlnAlaGlnAsnV 1038

1742 ACAACAGCTTGGATAGCAAAAAAGAAATTGCC......TACAACGGTTGG 1785
        ::::::|||:::   |||:::   |||            ||||||
1038 alAlaGlyValAspThrValLysSerSerAlaAsnThrLeuAsnGlyAla 1054

1786 TTTGGCGAG..............AAAGATACGACCAAAACGAACGGGCG 1820
     :::|||                    :::|||:::|||   ||||||:::
1055 MetGlyThrLeuArgAsnSerIleGlnAspAsnThrAlaThrAsnAsnGl 1071

1821 GCTCAACCTTGTTTACCAGCCCGCCGCAGAAGACCGCACCCTGCTGCTTT 1870
        |||      |||      |||:::|||                    |
1071 yGlnAsn......TyrLeuAspAlaThrGlu..................S 1080

1871 CCGGCGGAACAAATTTAAACGGCAACATCACGCAAACAAACGGCAAACTG 1920
     ||:::  ||||||  |||:::   :::   ::::::||||||   :::
1080 erAsnLysThrAsnTyrAsnAsnAlaValAspSerAlaAsnGlyValIle 1096

1921 TTTTCAGCGGCAGACCAACACCGCACGCCTACAATCATTTAAACGACCA 1970
         ::::::::|||:::         |||   |||   :::|||:::
1097 AsnAlaThrSerAsnProAsnMetAspAla...AsnAlaIleAsnGlnIl 1112

1971 TTGGTCGCAA............................AAAGAGGGCA 1990
        :::|||                                 :::::::
1112 eAlaThrGlnValThrSerThrLysAsnAlaLeuAspGlyThrHisAsnL 1129

1991 TTCCTCGCGGGGAAATCGTGTGGGACAAC......GACTGGATCAACCGC 2034
     ::   :::::::::   :::      |||      |||      ::::::
1129 euThrGlnAlaLysGlnThrAlaThrAsnAlaIleAspGlyAlaThrAsn 1145

2035 ACATTTAAAGCGGAAAACTTCCAAATTAAAGGCGGACAGGCGGTGGTTTC 2084
         ||||||::::::      :::|||   :::|||::::::   ::
1146 LeuAsnLysAlaGlnLysAspAlaLeuLys...AlaGlnValThrSerAl 1161

2085 CCGCAATGTTGCCAAAGTGAAAGGCGATTGGCATTTGAGCAATCACGCCC 2134
     :::::::||||||:::|||   :::      :::   :::|||:::   :
1161 aGlnArgValAlaAsnValThrSerIleGlnGlnThrAlaAsnGluLeuA 1178

2135 AAGCAGTTTTTGGTGTCGCACCGCAT......................... 2160
     ::::::::::|||         |||
1178 snThrAlaMetGlyGlnLeuGlnHisGlyIleAspAspGluAsnAlaThr 1194

2161 .........CAAAGCCACACAATCTGTACACGTTCGGACTGGACGGGTCT 2201
              |||::::::   :::   :::|||
1195 LysGlnThrGlnLysTyrArgAspAlaGluGlnSer.............. 1206

2202 GACAAATTGTGTCGAAAAAACCATTACCGACGATAAAGTGATTGCT.... 2247
                 :::||||||   |||:::   |||   |||
1207 .............LysLysThrAlaTyrAspGlnAlaValAlaAlaAlaL 1219

2248 ..TCATTGACTAAGACCGACATCAGCGGCAATGTCGATCTTGCC...... 2289
       ::::::  :::   :::   ::::::|||   |||   |||
1219 ysAlaIleLeuAsnLysGlnThrGlySerAsnSerAspLysAlaAlaVal 1235

2290 GATCACGCTCATTTAAATCTCACAGGGCTTGCCACACTCAACGGCAATCT 2339
     |||:::|||        ::::::|||:::
1236 AspArgAlaLeuGlnGlnValThrSerThrLysAsp.............. 1247

2340 TAGTGCAAATGGCGATACACGTTATACAGTCAGCCACAACGCCACC...C 2386
      :::   |||||||||::::::   :::   :::      |||:::   |
1248 .AlaLeuAsnGlyAspAlaLysLeuAlaGluAlaLysAlaAlaAlaLysG 1264

2387 AAAAC...GGCAACCTTAGCCTCGTGGGCAATGCCCAAGCAACA...... 2427
     |||||   |||:::|||:::   :::   |||||||||   |||
1264 lnAsnLeuGlyThrLeuAsnHisIleThrAsnAlaGlnArgThrAlaLeu 1280

2428 .........TTTAATCAAGCC...ACATTAAACGGCAACACATCGGCTTC 2465
              :::||||||||   |||::::::|||   ::::::::::
1281 GluGlyGlnIleAsnGlnAlaThrThrValAspGlyValAsnThrValLy 1297

2466 GGGCAATGCT.................................TCAT 2479
     :   ||||||                                   |||:
1297 sThrAsnAlaAsnThrLeuAspGlyAlaMetAsnSerLeuGlnGlySerI 1314
```

```
2480 TTAATCTAAGCGACCACGCCGTACAAAACGGCAGTCTGACGCTTTCCGGC 2529
     ::|||   :::|||   :::::::::|||   :::         :::
1314 leAsnAspLysAspAlaThrLeuArgAsnGlnAsnTyrLeuAspAlaAsp 1330

2530 AACGCTAAGGCAAACGTAAGCCATTCCGCACTC................. 2562
     ::::::|||   |||:::   :::|||:::
1331 GluSerLysArgAsnAlaTyrThrGlnAlaValThrAlaAlaGluGlyIl 1347

2563 ................AACGGTAATGTCTCCCTAGCCGAT.......... 2586
                     :::||||||:::|||   ||||||
1347 eLeuAsnLysGlnThrGlyGlyAsnThrSerLysAlaAspValAspAsnA 1364

2587 .......................AAGGCAGTATTCCAT...TTTGAAAGC 2610
                            ||||||:::::::::      |||:::
1364 laLeuAsnThrValThrArgAlaLysAlaAlaLeuAsnGlyAlaGluAsn 1380

2611 AGCCGC...............TTTACCGGACAAATCAGCGGC........ 2637
       |||                   |||:::   |||:::|||
1381 LeuArgAsnThrLysThrSerAlaThrAsnThrIleAsnGlyLeuProAs 1397

2638 .............GGCAAGGATACGGCATTACAC................ 2658
                  ||||||:::     |||
1397 nLeuThrGlnLeuGlnLysAspAsnLeuLysHisGlnValGluGlnAlaG 1414

2659 .......................TTAAAAGACAGCGAATGGACGCTGCCG 2685
                            :::||||||:::        |||||
1414 lnAsnValAlaGlyValAsnGlyValLysAspLysGlyAsnThrLeu... 1429

2686 TCAGGCACGGAATTAGGCAATTTA......AACCTTGACAACGCCACCAT 2729
        :::|||   :::|||   |||        :::::::::|||   :::
1430 ...AsnThrAlaMetGlyAlaLeuArgThrSerIleGlnAsnAspAsnTh 1445

2730 TACACTCAATTCCGCCTATCGCCACGATGCGGCAGGGGCGCAAACCGGCA 2779
      |||   :::|||               ||||||:::   ::::::   ::::
1445 rThrLysThrSerGlnAsnTyrLeuAspAlaSerAspSerAsnLysAsnA 1462

2780 GTGCGACAGATGCGCCGCGCCGCCGTTCGCGCCGTTCGCGCCGTTCCCTA 2829
     ::   :::   |||                   ::::::   :::::::::
1462 snTyrAsnThrAlaVal..............AsnAsnAlaAsnGlyVal 1473

2830 TTATCCGTTACACCGCCAACTTCGGTAGAATCCCGTTTCAACACGCTGAC 2879
     ::::::||||||            ::::::::::::::          ::
1474 IleAsnValThrAsnAsnProAsnMetAspAlaAsn............Al 1486

2880 GGTAAACGGCAAATTGAACGGTCAGGGAACATTCCGCTTTATGTCGGAAC 2929
     ::::||||||      |||      :::|||   :::      :::    |
1486 aIleAsnGlyMetAlaAsnGlnValAsnThrThrLys......AlaAlaL 1501

2930 TCTTCGGCTACCGCAGCGACAAATTGAAGCTGGCGGAAAGTTCCGAAGGC 2979
      ||   |||   ::::::            ||||||:::::::::
1501 euAsnGlyAlaGlnAsn............LeuAlaGlnAlaLys...... 1511

2980 ACTTACACCTTGGCGGTCAACAATACCGGCAACGAACCTGCA........ 3021
           |||   |||   ||||||   :::|||      :::
1512 ......ThrThrAlaLysAsnAsnLeuAsnAsnLeuThrSerIleAsnAs 1526

3022 ..........................AGCCTCGAACAATTGACGGTAGTGG 3046
                               |||::::::      |||:::|||
1526 nAlaGlnLysAspAlaLeuThrArgSerIleAspGlyAlaThrThrValA 1543

3047 AAGGAAAAGACAACAAACCGCTGTCC............GAAAACCTTAAT 3084
        |||   :::::::::      :::            :::|||
1543 laGlyValAsnGlnGluThrAlaLysAlaThrGluLeuAsnAsnAlaMet 1559

3085 TTCACCCTGCAAAACGAACACGTCGATGCCGGCGCGTGGCGTTACCAACT 3134
       :::|||||||||
1560 HisSerLeuGlnAsn.................................. 1564

3135 CATCCGCAAAGACGGCGAGTTCCGCCTGCATAATCCGGTCAAAGAACAAG 3184
                               :::::::::   ::::::   ::::
1565 ......................GlyIleAsnAspGluThrGlnThrLysG 1574

3185 AGCTTTCCGACAAACTCGGCAAGGCAGAAGCCAAAAAACAGGCGGAAAAA 3234
     ::    :::       |||       :::      :::|||||||:::|||
1574 lnThrGlnLysTyrLeuAspAlaGlyProSerLysLysSerAlaTyrAsp 1590

3235 GACAACGCGCAAAGCCTTGACGCGCTGATTGCGGCCGGGCGCGATGCCGT 3284
     :::   :::::::::         |||::::::::::
1591 GlnAlaValAsnAlaAlaLysAlaIleLeuThr................. 1601

3285 CGAAAAGACAGAAAGCGTTGCCGAACCGGCCCGGCAGGCAGGCGGGGAAA 3334
                                         :::|||:::|||:::|
1602 .................................LysAlaSerGlyGlnA 1607

3335 ATGTCGGCATTATGCAGGCGGAGGAAGAGAAAAAACGGGTGCAGGCGGAT 3384
      |||||            :::|||:::         ::::::|||::::::
1607 snValAspLysAlaAlaValGluGlnAlaLeuGlnAsnValAsnSerThr 1623

3385 AAAGACACCGCCTTG.........GCGAAACAGCGCGAAGCG........ 3417
      |||   |||||||||            ||||||   :::||||||
1624 Lys...ThrAlaLeuAsnGlyAspAlaLysLeuAsnGluAlaLysAlaAl 1639

3418 .......GAAACCCGGCCGGCTACCACCGCCTTCCCCCGCGCCCGCCGCG 3460
            :::|||      :::      |||   :::   :::|||:::|||:
1639 aAlaLysGlnThrLeuGlyThrLeuThrHisIleAsnAsnAlaGlnArgT 1656

3461 CCCGCCGGGATTTGCCGCAACTGCAACCCCAACCGCAGCCCCAACCGCAG 3510
     ::      |||
1656 hrAlaLeuAsp...................................... 1659

3511 CGCGACCTGATCAGCCGTTATGCCAATAGCGGTTTGAGTGAATTTTCCGC 3560
        :::   |||::::::   :::|||
1660 ...AsnGluIleThrGlnAlaThrAsn.................ValGl 1669

3561 CACGCTCAACAGCGTTTTCGCCGTACAGGACGAATTAGACCGCGTATTTG 3610
        :::|||:::|||   |||         ::::::||||||   :::::::
1669 uGlyValAsnThrValLysAlaLysAlaGlnGlnLeuAspGlyAlaMetG 1686

3611 CCGAAGACCGCCGCAACGCCGTTTGGACAAGCGGCATCCGGGACACCAAA 3660
     :::::                         ::::::|||||||||       ::
1686 lyGln.................LeuGluThrSerIleArgAsp.LysAs 1696

3661 CACTACCGTTCGCAAGATTTCCGCGCCTACCGCCAACAAACCGACCTGCG 3710
     :||||||:::::::::   :::         |||:::            |
1696 pThrThrLeuGlnSerGlnAsnTyrGlnAspAlaAspAsp.........A 1710

3711 CCAAATCGGTATGCAGAAAAACCTCGGCAGCGGGCGCGTCGGCATCCTGT 3760
      |||||   ::::::   :::   ::::::   :::|||::::::
1710 laLysArgThrAlaTyrSerGlnAlaValAsnAlaAlaAlaThrIleLeu 1726

3761 TTTCGCACAACCGGACCGAAAACACCTTCG.................... 3790
        :::|||:::|||   :::|||||||:::
1727 AsnLysThrAlaGlyGlyAsnThrProLysAlaAspValGluArgAlaMe 1743

3791 .........ACGACGGCATCGGCAACTCGGCACGGCTTGCCCACGGCGC 3830
                |||   |||:::::::::   :::|||:::   :::
1743 tGlnAlaValThrGlnAlaAsnThrAlaLeuAsnGlyIleGlnAsnLeuG 1760

3831 CGTTTTCGGGCAATACGGCATCGACAGGTTCTACATCGGCATCA...... 3874
        :::   ::::::|||:::|||:::   |||:::||||||
1760 luArgAlaLysGlnAlaAlaAsnThrAlaIleThrAsnAlaSerAspLeu 1776

3875 ......................................GCGCGGGCGCG 3885
                                           :::   :::||
1777 AsnThrLysGlnLysGluAlaLeuLysAlaGlnValThrSerAlaGlyAr 1793

3886 GGTTTAGCAGCGGCAGCCTTTCAG...ACGGCATCGGAGGCAAAATCCG 3932
      ||||   ||||||   :::   :::   ||||||:::|||   ::::::|
1793 gValSerAlaAlaAsnGlyValGluHisThrAlaThrGluLeuAsnThrA 1810

3933 CCGCCGCGTGCTGCATTACGGCATTCAGGCAC.........GATACCGCG 3973
      ||           :::||||||::::::|||         |||   |||
1810 la...........MetThrAlaLeuLysHisAlaIleAlaAspLysAla 1822

3974 CCGGTTTCGGCGGATTCGGCATCGAACCGCACATCGGCGCAACGCGCTAT 4023
        ::::::|||::::::            :::::::::|||::::::
```

```
1823 GluThrLysAlaSerGlyAsnTyrValAsnAlaAspAlaAsnLysArgGl 1839

4024 TTCGTCCAAAAAGCGGATTACCGCTACGAAAACGTCAATATCG......C 4067
       :::      ::::::::||||||:::::::::   :::|||      |
1839 nAlaTyrAspGluLysValThrAlaAlaGluAsnIleValSerGlyThrP 1856

4068 CACCCCCGGCCTTGCATTCAACCGCTACCGCGCGGCATTAAGGCAGATT 4117
      ||   |||:::|||       :::   :::::::::|||      |||:::
1856 roThrProThrLeuThrProSerAspValThrAsnAlaAlaThrGlnVal 1872

4118 ATTCATTCA 4126
        ::::::
1873 ThrAsnAla 1875
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAU37017

```
seq_documentation_block:
ID   AAU37017 standard; Protein; 5795 AA.
XX
AC   AAU37017;
XX
DT   14-FEB-2002  (first entry)
XX
DE   Staphylococcus aureus cellular proliferation protein #1187.
XX
KW   Antisense; prokaryotic cellular proliferation protein;
KW   antibiotic; antibacterial; drug design.
XX
OS   Staphylococcus aureus.
XX
PN   WO200170955-A2.
XX
PD   27-SEP-2001.
XX
PF   21-MAR-2001; 2001WO-US09180.
XX
PR   21-MAR-2000; 2000US-191078P.
PR   23-MAY-2000; 2000US-206848P.
PR   26-MAY-2000; 2000US-207727P.
PR   23-OCT-2000; 2000US-242578P.
PR   27-NOV-2000; 2000US-253625P.
PR   22-DEC-2000; 2000US-257931P.
PR   16-FEB-2001; 2001US-269308P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Haselbeck R,  Ohlsen KL,  Zyskind JW,  Wall D,  Trawick JD,  Carr GJ;
PI   Yamamoto RT,  Xu HH;
XX
DR   WPI; 2001-611495/70.
DR   N-PSDB; AAS54876.
XX
PT   New polynucleotides for the identification and development of
PT   antibiotics, comprise sequences of antisense nucleic acids -
XX
PS   Example 3; Seq ID No 12610; 511pp; English.
XX
CC   The invention relates to antisense inhibitors of genes essential to
CC   prokaryotic cellular proliferation, their use in identifying the
CC   genes, their use in the discovery of novel antibiotics, the essential
CC   genes themselves and the encoded proteins. The prokaryotes used are
CC   Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC   pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC   invention is also useful for the identification of potential new targets
CC   for antibiotic development. The antisense nucleic acids can also be used
CC   to identify proteins used in proliferation, to express these proteins,
CC   and to obtain antibodies capable of binding to the expressed proteins.
CC   The proteins can be used to screen compounds in rational drug discovery
CC   programmes. The antisense nucleic acid sequence is also useful to screen
CC   for homologous nucleic acids which are required for cell proliferation in
CC   a wide variety of organisms. The present sequence represents an
CC   essential prokaryotic cellular proliferation protein.
CC   Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic
CC   format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   5795 AA;
```

```
alignment_scores:
           Quality:  264.50            Length:   1388
             Ratio:  0.387               Gaps:     69
 Percent Similarity:  49.207   Percent Identity: 20.605

alignment_block:
 US-09-303-518D-649 x AAU37017   ..

 Align seg 1/1  to: AAU37017  from: 1  to: 5795

   505 ACAGATGCAGAACCTGTTGAAATGACCAGTTAT...........ATGGA 542
        |||||||||:::|||   ::::::::::|||              ::
  2974 ThrAspAlaSerProAsnAsnGlnSerThrTyrAsnSerAlaValSerAs 2990

   543 TGGGCGGAAATATATCGATCAAAATAATTACCCTGACCGTGTTCGTATTG 592
        :::::::      |||:::|||:::|||   |||
  2990 nAlaLysGlyIleIleAsnGlnThrAsnAsnPro................ 3001

   593 GGGCAGGCAGGCAATATTGGCGATCTGATGAAGATGAGCCCAATAACCGC 642

  3001 .................................................. 3001

   643 GAAAGTTCATATCATATTGCAAGTGCGTATTCTTGGCTCGTTGGTGGCAA 692
                         :::         |||   ::::::   :::::
  3002 .................ThrMetAspThrSerAlaIleThrGlnAlaTh 3012

   693 TACCTTTGCACAAAATGGATCAGGTGGTGGCACAGTCAACTTAGGTAGTG 742
        :|||   ::::::|||:::::::::|||           :::|||   |||:::|
  3012 rThrGlnValAsnAsnAlaLysAsnGly......LeuAsn...GlyAlaG 3026

   743 AAAAAATTAAACATAGCCCATATGGTTTTTACCAACAGGAGGCTCATTT 792
        ||::::::::::: |||:::   :::       |||        :::||
  3026 luAsnLeuArgAsn.AlaGlnAsnThrAlaLysGln.........AsnLe 3039

   793 GGCGACAGTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCAAAAGTG 842
        |   |||::::::|||               :::   ::::::::::::
  3039 uAsnThrLeuSerHis..............LeuThrAsnAsnGlnLysS 3051

   843 GTTAATTAATGGGGTATTGCAAACGGGCAACCCCTATATAGGAAAAAGCA 892
        ::  :::   :::             ||||||      :::   |||   :::
  3051 erAlaIleSerSerGlnIleAspArgAlaGlyHisValSerGluValThr 3067

   893 ATGGCTTCCAGCTGGTTCGTAAAGATTGGTTCTATGATGAAATCTTTGCT 942
           |||::::::         :::::::::::::::|||    :::|||
  3068 AlaAlaLysAsnAlaAlaThrGluLeuAsnThrGlnMetGlyAsnLeuGl 3084

   943 GGAGATACCCATTCAGTATTCTACGAACCACGTCAAAATGGGAAATACTC 992
        :::             ::::::  |||   ||||||:::|||       ::
  3084 uGln......AlaIleHisAspGlnAsnThrValLysGlnGly..ValAs 3098

   993 TTTTAACGACGATAATAATGGCACAGGAAAAATCAATGCCAAACATGAAC 1042
        :|||:::|||   :::                           |||   ::::
  3098 nPheThrAspAlaAsp......................LysAlaLysA 3107

  1043 ACAATTCTCTGCCTAATAGATTAAAAACACGAACCGTTCAATTGTTTAAT 1092
        :::::::      |||   ::::::          :::   |||   :::
  3107 rgAspAlaTyrThrAsnAlaValSerArgAlaGluThrIleLeuAsnLys 3123

  1093 GTTTCTTTATCCGAGACAGCAAGAGAACCTGTTTATCATGCTGCAGGTGG 1142
        ::::::   ::::::|||::::::::::   |||      |||      ::
  3124 ThrGlnGlyAlaAsnThrSerLysGlnAspValGluAlaAlaIleGlnAs 3140

  1143 TGTCAACAGTTATCGACCCAGACTGAATAATGGAGAAAATATTTCCTTTA 1192
        :|||:::|||   :::      ||||||:::   :::|||:::::
  3140 nValThrSerAlaLysAsnAlaLeuAsnGlyAspGlnAsnValThrAsnA 3157
```

```
1193 TT...GACGAAGGAAAAGGCGAATTGATACTTACCAGCAACATCAATCAA 1239
           :::    :::|||      |||         ::::::|||||||:::
3157 laLysAsnThrAlaLysHisAlaLeuAsnAsnLeuThrSerIleAsnAsn 3173

1240 GGTGCTGGAGGATTATATTTCCAAGGAGATTTTACGGTCTCGCCTGAAAA 1289
     :::                     :::   |||:::|||::::::   :::::
3174 AlaGln..............LysArgAspLeuThrThrLysIleAspGl 3185

1290 TAACGAAACTTGGCAAGGCGCGGGCGTTCATATCAGTGAAGACAGTACCG 1339
     :      |||      |||
3185 nAlaThrThrValAlaGly............................. 3191

1340 TTACTTGGAAAGTAAACGGCGTGGCAAACGACCGCCTGTCCAAAATCGGC 1389
                |||::::::|||:::|||
3192 ..........ValGluAlaValSerAsn.................... 3197

1390 AAAGGCACGCTGCACGTTCAAGCCAAAGGGGAAAACCAAGGCTCGATCAG 1439
       ||||||          |||   ::::::   |||::::::|||:::
3198 ThrGlyThrGlnLeuAsnThrAlaMetAlaAsnLeuGlnAsnGlyIleAs 3214

1440 CGTGGGCGACGGTACAGTCATTTTGGATCAGCAGGCAGAC...GATAAAG 1486
     :          :::|||:::      ::::::      |||   |||:::
3214 nAspLysAlaAsnThrLeuAlaSerGluAsnTyrHisAspAlaAspSerA 3231

1487 GCAAAAAACAAGCCTTTAGTGAAATCGGCTTGGTCAGCGGCAGGGGTACG 1536
       ||||||   |||:::::::::
3231 spLysLysThrAlaTyrThrGln......................... 3238

1537 GTGCAACTGAATGCCGATAATCAGTTCAACCCCGACAAACTCTATTTCGG 1586
     :::        |||||||:::|||   :::|||   ::::::
3239 AlaValThrAsnAlaGluAsnIleLeuAsnLysAsnSer........... 3251

1587 CTTTCGCGGCGGACGTTTGGATTTAAACGGGCATTCGCTTTCGTTCCACC 1636
            |||::::::||||||   :::   :::   ::::::    :
3252 .......GlySerAsnLeuAspLysAlaAlaValGluAsnAlaLeuSerG 3266

1637 GTATTCAAAATACCGATGAAGGGGCGATGATTGTCAACCACAATCAAGAC 1686
     :::::   |||:::   :::||||||:::       ||||||||   :::
3266 lnValThrAsnAla...LysGlyAlaLeuAsnGlyAsnHisAsnLeuGlu 3281

1687 AAA.........GAATCCACCGTTACCATTACAGGCAATAAAGATATTGC 1727
     :::           :::::::::::||||||:::|||   :::   :::::
3282 GlnAlaLysSerAsnAlaAsnThrThrIleAsnGlyLeuGlnHisLeuTh 3298

1728 TACA..................................ACCGGCAATA 1741
     :|||                                  :::   |||
3298 rThrAlaGlnLysAspLysLeuLysGlnGlnValGlnGlnAlaGlnAsnV 3315

1742 ACAACAGCTTGGATAGCAAAAAAGAAATTGCC......TACAACGGTTGG 1785
         ::::::|||:::   |||:::   |||       ||||||
3315 alAlaGlyValAspThrValLysSerSerAlaAsnThrLeuAsnGlyAla 3331

1786 TTTGGCGAG..............AAAGATACGACCAAAACGAACGGGCG 1820
     :::|||                  :::|||:::|||   |||||||:::
3332 MetGlyThrLeuArgAsnSerIleGlnAspAsnThrAlaThrAsnAsnGl 3348

1821 GCTCAACCTTGTTTACCAGCCCGCCGCAGAAGACCGCACCCTGCTGCTTT 1870
        |||      |||      |||:::|||                    |
3348 yGlnAsn......TyrLeuAspAlaThrGlu.................S 3357

1871 CCGGCGGAACAAATTTAAACGGCAACATCACGCAAACAAACGGCAAACTG 1920
     ||:::   ||||||   |||:::   :::   ::::::||||||   :::
3357 erAsnLysThrAsnTyrAsnAsnAlaValAspSerAlaAsnGlyValIle 3373

1921 TTTTCAGCGGCAGACCAACACCGCACGCCTACAATCATTTAAACGACCA 1970
           :::::::::|||:::      |||   |||   :::|||:::
3374 AsnAlaThrSerAsnProAsnMetAspAla...AsnAlaIleAsnGlnIl 3389

1971 TTGGTCGCAA............................AAAGAGGGCA 1990
        :::|||                              :::::::
3389 eAlaThrGlnValThrSerThrLysAsnAlaLeuAspGlyThrHisAsnL 3406

1991 TTCCTCGCGGGGAAATCGTGTGGGACAAC......GACTGGATCAACCGC 2034
     ::   :::::::::   :::       |||       |||       ::::::
3406 euThrGlnAlaLysGlnThrAlaThrAsnAlaIleAspGlyAlaThrAsn 3422

2035 ACATTTAAAGCGGAAAACTTCCAAATTAAAGGCGGACAGGCGGTGGTTTC 2084
        ||||||::::::         :::|||   :::|||::::::   ::
3423 LeuAsnLysAlaGlnLysAspAlaLeuLys...AlaGlnValThrSerAl 3438

2085 CCGCAATGTTGCCAAAGTGAAAGGCGATTGGCATTTGAGCAATCACGCCC 2134
     ::::::|||||||:::|||   :::         :::   :::|||:::   :
3438 aGlnArgValAlaAsnValThrSerIleGlnGlnThrAlaAsnGluLeuA 3455

2135 AAGCAGTTTTGGTGTCGCACCGCAT........................ 2160
     ::::::::::|||          |||
3455 snThrAlaMetGlyGlnLeuGlnHisGlyIleAspAspGluAsnAlaThr 3471

2161 .........CAAAGCCACACAATCTGTACACGTTCGGACTGGACGGGTCT 2201
              |||::::::      :::   :::|||
3472 LysGlnThrGlnLysTyrArgAspAlaGluGlnSer............. 3483

2202 GACAAATTGTGTCGAAAAAACCATTACCGACGATAAAGTGATTGCT.... 2247
                 :::||||||   |||:::   |||   |||
3484 ............LysLysThrAlaTyrAspGlnAlaValAlaAlaAlaL 3496

2248 ..TCATTGACTAAGACCGACATCAGCGGCAATGTCGATCTTGCC...... 2289
       ::::::   :::   :::   ::::::|||   |||   |||
3496 ysAlaIleLeuAsnLysGlnThrGlySerAsnSerAspLysAlaAlaVal 3512

2290 GATCACGCTCATTTAAATCTCACAGGGCTTGCCACACTCAACGGCAATCT 2339
     |||:::|||         ::::::|||:::
3513 AspArgAlaLeuGlnGlnValThrSerThrLysAsp............. 3524

2340 TAGTGCAAATGGCGATACACGTTATACAGTCAGCCACAACGCCACC...C 2386
      :::   |||||||||::::::   :::   :::         |||:::   |
3525 .AlaLeuAsnGlyAspAlaLysLeuAlaGluAlaLysAlaAlaAlaLysG 3541

2387 AAAAC...GGCAACCTTAGCCTCGTGGGCAATGCCCAAGCAACA...... 2427
     |||||   |||:::|||:::   :::   |||||||||   |||
3541 lnAsnLeuGlyThrLeuAsnHisIleThrAsnAlaGlnArgThrAlaLeu 3557

2428 .........TTTAATCAAGCC...ACATTAAACGGCAACACATCGGCTTC 2465
              :::|||||||||   |||::::::|||   :::::::::::
3558 GluGlyGlnIleAsnGlnAlaThrThrValAspGlyValAsnThrValLy 3574

2466 GGGCAATGCT..................................TCAT 2479
     :    ||||||                                   |||:
3574 sThrAsnAlaAsnThrLeuAspGlyAlaMetAsnSerLeuGlnGlySerI 3591

2480 TTAATCTAAGCGACCACGCCGTACAAAACGGCAGTCTGACGCTTTCCGGC 2529
     ::|||   :::|||   :::::::::|||   :::         :::
3591 leAsnAspLysAspAlaThrLeuArgAsnGlnAsnTyrLeuAspAlaAsp 3607

2530 AACGCTAAGGCAAACGTAAGCCATTCCGCACTC................ 2562
     ::::::|||   |||:::      :::|||:::
3608 GluSerLysArgAsnAlaTyrThrGlnAlaValThrAlaAlaGluGlyIl 3624

2563 ................AACGGTAATGTCTCCCTAGCCGAT.......... 2586
                     :::|||||||:::|||   ||||||
3624 eLeuAsnLysGlnThrGlyGlyAsnThrSerLysAlaAspValAspAsnA 3641

2587 .......................AAGGCAGTATTCCAT...TTTGAAAGC 2610
                            ||||||::::::::::      |||:::
3641 laLeuAsnThrValThrArgAlaLysAlaAlaLeuAsnGlyAlaGluAsn 3657

2611 AGCCGC...............TTTACCGGACAAATCAGCGGC........ 2637
        |||                |||:::   |||:::|||
3658 LeuArgAsnThrLysThrSerAlaThrAsnThrIleAsnGlyLeuProAs 3674

2638 ............GGCAAGGATACGGCATTACAC................ 2658
                 ||||||:::      |||
3674 nLeuThrGlnLeuGlnLysAspAsnLeuLysHisGlnValGluGlnAlaG 3691

2659 .......................TTAAAAGACAGCGAATGGACGCTGCCG 2685
                            :::||||||:::      ||||||
```

```
3691 lnAsnValAlaGlyValAsnGlyValLysAspLysGlyAsnThrLeu... 3706

2686 TCAGGCACGGAATTAGGCAATTTA......AACCTTGACAACGCCACCAT 2729
        :::|||   :::|||  |||      :::::::::|||   :::
3707 ...AsnThrAlaMetGlyAlaLeuArgThrSerIleGlnAsnAspAsnTh 3722

2730 TACACTCAATTCCGCCTATCGCCACGATGCGGCAGGGGCGCAAACCGGCA 2779
       |||   :::|||            ||||||:::   ::::::   ::::
3722 rThrLysThrSerGlnAsnTyrLeuAspAlaSerAspSerAsnLysAsnA 3739

2780 GTGCGACAGATGCGCCGCGCCGCCGTTCGCGCCGTTCGCGCCGTTCCCTA 2829
     ::   :::   |||                 ::::::   :::::::::
3739 snTyrAsnThrAlaVal..............AsnAsnAlaAsnGlyVal 3750

2830 TTATCCGTTACACCGCCAACTTCGGTAGAATCCCGTTTCAACACGCTGAC 2879
     ::::::||||||       ::::::::::::::                ::
3751 IleAsnValThrAsnAsnProAsnMetAspAlaAsn............Al 3763

2880 GGTAAACGGCAAATTGAACGGTCAGGGAACATTCCGCTTTATGTCGGAAC 2929
     ::::||||||     |||     :::|||   :::      :::   |
3763 aIleAsnGlyMetAlaAsnGlnValAsnThrThrLys......AlaAlaL 3778

2930 TCTTCGGCTACCGCAGCGACAAATTGAAGCTGGCGGAAAGTTCCGAAGGC 2979
     ||   |||   ::::::              ||||||:::::::::
3778 euAsnGlyAlaGlnAsn...........LeuAlaGlnAlaLys...... 3788

2980 ACTTACACCTTGGCGGTCAACAATACCGGCAACGAACCTGCA........ 3021
           |||   |||   ||||||   :::|||      :::
3789 ......ThrThrAlaLysAsnAsnLeuAsnAsnLeuThrSerIleAsnAs 3803

3022 ........................AGCCTCGAACAATTGACGGTAGTGG 3046
                              |||:::::::      |||:::|||
3803 nAlaGlnLysAspAlaLeuThrArgSerIleAspGlyAlaThrThrValA 3820

3047 AAGGAAAAGACAACAAACCGCTGTCC...........GAAAACCTTAAT 3084
       |||   :::::::::      :::             :::|||
3820 laGlyValAsnGlnGluThrAlaLysAlaThrGluLeuAsnAsnAlaMet 3836

3085 TTCACCCTGCAAAACGAACACGTCGATGCCGGCGCGTGGCGTTACCAACT 3134
        :::|||||||||
3837 HisSerLeuGlnAsn................................. 3841

3135 CATCCGCAAAGACGGCGAGTTCCGCCTGCATAATCCGGTCAAAGAACAAG 3184
                             :::::::::   ::::::   ::::
3842 .....................GlyIleAsnAspGluThrGlnThrLysG 3851

3185 AGCTTTCCGACAAACTCGGCAAGGCAGAAGCCAAAAAACAGGCGGAAAAA 3234
     ::   :::         |||      :::   :::||||||:::|||
3851 lnThrGlnLysTyrLeuAspAlaGlyProSerLysLysSerAlaTyrAsp 3867

3235 GACAACGCGCAAAGCCTTGACGCGCTGATTGCGGCCGGGCGCGATGCCGT 3284
     :::   :::::::::         |||:::::::::
3868 GlnAlaValAsnAlaAlaLysAlaIleLeuThr................ 3878

3285 CGAAAAGACAGAAAGCGTTGCCGAACCGGCCCGGCAGGCAGGCGGGGAAA 3334
                                 :::|||:::|||:::|
3879 ...............................LysAlaSerGlyGlnA 3884

3335 ATGTCGGCATTATGCAGGCGGAGGAAGAGAAAAAACGGGTGCAGGCGGAT 3384
     |||||          :::|||:::      ::::::|||::::::
3884 snValAspLysAlaAlaValGluGlnAlaLeuGlnAsnValAsnSerThr 3900

3385 AAAGACACCGCCTTG.........GCGAAACAGCGCGAAGCG........ 3417
     |||   |||||||||          ||||||   :::||||||
3901 Lys...ThrAlaLeuAsnGlyAspAlaLysLeuAsnGluAlaLysAlaAl 3916

3418 .......GAAACCCGGCCGGCTACCACCGCCTTCCCCCGCGCCCGCCGCG 3460
           :::|||      :::   |||   :::   :::|||:::|||:
3916 aAlaLysGlnThrLeuGlyThrLeuThrHisIleAsnAsnAlaGlnArgT 3933

3461 CCCGCCGGGATTTGCCGCAACTGCAACCCCAACCGCAGCCCCAACCGCAG 3510
     ::      |||
3933 hrAlaLeuAsp..................................... 3936

3511 CGCGACCTGATCAGCCGTTATGCCAATAGCGGTTTGAGTGAATTTTCCGC 3560
        :::   |||::::::   :::|||
3937 ...AsnGluIleThrGlnAlaThrAsn.................ValGl 3946

3561 CACGCTCAACAGCGTTTTCGCCGTACAGGACGAATTAGACCGCGTATTTG 3610
        :::|||:::|||   |||      ::::::||||||   ::::::
3946 uGlyValAsnThrValLysAlaLysAlaGlnGlnLeuAspGlyAlaMetG 3963

3611 CCGAAGACCGCCGCAACGCCGTTTGGACAAGCGGCATCCGGGACACCAAA 3660
     :::::                              ::::::|||||||||   ::
3963 lyGln.................LeuGluThrSerIleArgAsp.LysAs 3973

3661 CACTACCGTTCGCAAGATTTCCGCGCCTACCGCCAACAAACCGACCTGCG 3710
     :||||||:::::::::   :::         |||:::           |
3973 pThrThrLeuGlnSerGlnAsnTyrGlnAspAlaAspAsp.........A 3987

3711 CCAAATCGGTATGCAGAAAAACCTCGGCAGCGGGCGCGTCGGCATCCTGT 3760
     |||||   ::::::   :::   ::::::   :::|||::::::
3987 laLysArgThrAlaTyrSerGlnAlaValAsnAlaAlaThrIleLeu 4003

3761 TTTCGCACAACCGGACCGAAAACACCTTCG.................... 3790
        :::|||:::|||   :::||||||:::
4004 AsnLysThrAlaGlyGlyAsnThrProLysAlaAspValGluArgAlaMe 4020

3791 .........ACGACGGCATCGGCAACTCGGCACGGCTTGCCCACGGCGC 3830
              |||   |||:::::::::   :::|||:::   :::
4020 tGlnAlaValThrGlnAlaAsnThrAlaLeuAsnGlyIleGlnAsnLeuG 4037

3831 CGTTTTCGGGCAATACGGCATCGACAGGTTCTACATCGGCATCA...... 3874
        :::   ::::::|||:::|||:::   |||:::||||||
4037 luArgAlaLysGlnAlaAlaAsnThrAlaIleThrAsnAlaSerAspLeu 4053

3875 .....................................GCGCGGGCGCG 3885
                                           :::   :::||
4054 AsnThrLysGlnLysGluAlaLeuLysAlaGlnValThrSerAlaGlyAr 4070

3886 GGTTTTAGCAGCGGCAGCCTTTCAG...ACGGCATCGGAGGCAAAATCCG 3932
     ||||   ||||||   :::   :::   ||||||:::|||   ::::::|
4070 gValSerAlaAlaAsnGlyValGluHisThrAlaThrGluLeuAsnThrA 4087

3933 CCGCCGCGTGCTGCATTACGGCATTCAGGCAC.........GATACCGCG 3973
     ||               :::||||||::::::|||         |||   |||
4087 la...........MetThrAlaLeuLysHisAlaIleAlaAspLysAla 4099

3974 CCGGTTTCGGCGGATTCGGCATCGAACCGCACATCGGCGCAACGCGCTAT 4023
           ::::::|||::::::      :::::::::|||::::::
4100 GluThrLysAlaSerGlyAsnTyrValAsnAlaAspAlaAsnLysArgGl 4116

4024 TTCGTCCAAAAAGCGGATTACCGCTACGAAAACGTCAATATCG......C 4067
      :::      :::::::::||||||:::::::::   :::|||       |
4116 nAlaTyrAspGluLysValThrAlaAlaGluAsnIleValSerGlyThrP 4133

4068 CACCCCCGGCCTTGCATTCAACCGCTACCGCGCGGGCATTAAGGCAGATT 4117
     ||   |||:::|||         :::   :::::::::|||      |||:::
4133 roThrProThrLeuThrProSerAspValThrAsnAlaAlaThrGlnVal 4149

4118 ATTCATTCA 4126
       ::::::
4150 ThrAsnAla 4152
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAU37403

```
seq_documentation_block:
ID   AAU37403 standard; Protein; 6281 AA.
XX
AC   AAU37403;
XX
DT   14-FEB-2002  (first entry)
XX
DE   Staphylococcus aureus cellular proliferation protein #1573.
XX
KW   Antisense; prokaryotic cellular proliferation protein;
```

```
KW   antibiotic; antibacterial; drug design.
XX
OS   Staphylococcus aureus.
XX
PN   WO200170955-A2.
XX
PD   27-SEP-2001.
XX
PF   21-MAR-2001; 2001WO-US09180.
XX
PR   21-MAR-2000; 2000US-191078P.
PR   23-MAY-2000; 2000US-206848P.
PR   26-MAY-2000; 2000US-207727P.
PR   23-OCT-2000; 2000US-242578P.
PR   27-NOV-2000; 2000US-253625P.
PR   22-DEC-2000; 2000US-257931P.
PR   16-FEB-2001; 2001US-269308P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Haselbeck R,  Ohlsen KL,  Zyskind JW,  Wall D,  Trawick JD,  Carr GJ;
PI   Yamamoto RT,  Xu HH;
XX
DR   WPI; 2001-611495/70.
DR   N-PSDB; AAS55262.
XX
PT   New polynucleotides for the identification and development of
PT   antibiotics, comprise sequences of antisense nucleic acids -
XX
PS   Example 3; Seq ID No 12996; 511pp; English.
XX
CC   The invention relates to antisense inhibitors of genes essential to
CC   prokaryotic cellular proliferation, their use in identifying the
CC   genes, their use in the discovery of novel antibiotics, the essential
CC   genes themselves and the encoded proteins. The prokaryotes used are
CC   Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC   pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC   invention is also useful for the identification of potential new targets
CC   for antibiotic development. The antisense nucleic acids can also be used
CC   to identify proteins used in proliferation, to express these proteins,
CC   and to obtain antibodies capable of binding to the expressed proteins.
CC   The proteins can be used to screen compounds in rational drug discovery
CC   programmes. The antisense nucleic acid sequence is also useful to screen
CC   for homologous nucleic acids which are required for cell proliferation in
CC   a wide variety of organisms. The present sequence represents an
CC   essential prokaryotic cellular proliferation protein.
CC   Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic
CC   format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   6281 AA;


alignment_scores:
           Quality:  262.50            Length:   1488
             Ratio:  0.382               Gaps:     66
 Percent Similarity:  46.237   Percent Identity: 19.825

alignment_block:
 US-09-303-518D-649 x AAU37403   ..

 Align seg 1/1  to: AAU37403  from: 1  to: 6281

    307 GCATTGGTGGGCGATCAATATATTGTGAGCGTGGCACATAACGGCGGCTA 356
         ||||||  |||||||||  :::::::::  |||  |||::::::
    217 AlaLeuAsnGlyAspGlnAsnValThrAsn...AlaLysAsnAlaAlaLy 232

    357 TAACAACGTTGATTTGGTGCGGAAGGAAGAAATCCCGATCAACATCGTT 406
           |||  ::::::      ::::::  :::|||  :::::::::   :
    232 sAsnAlaLeuAsnAsnLeuThrSerIleAsnAsnAlaGlnLysArgAspL 249

    407 TTACTTATAAAATTGTGAAACGGAATAATTATAAAGCAGGGACTAAAGGC 456
        ::|||  ||||||  :::  ::::::
    249 euThrThrLysIleAspGlnAlaThrThr..................... 258

    457 CATCCTTATGGCGGCGATTATCATATGCCGCGTTTGCATAAATTTGTCAC 506
                                                   |||::
    259 ...............................................ValAl 260

    507 AGATGCAGAACCTGTTGAAATGACCAGTTATATGGATGGGCGGAAATATA 556
        :  :::|||  |||:::  ||||||  :::   :::
    260 aGlyValGluAlaValSerAsnThrSerThrGlnLeuAsnThrAlaMetA 277

    557 TCGAT...CAAAATAATTACCCTGACCGTGTTCGTATTGGGGCAGGCAGG 603
          :::  ||||||:::   |||:::::::::   |||::::::
    277 laAsnLeuGlnAsnGlyIleAsnAspLysThrAsnThrLeuAlaSerGlu 293

    604 CAATATTGGCGATCTGATGAAGATGAGCCCAATAACCGCGAAAGTTCATA 653
        :::|||    :::|||:::|||:::
    294 AsnTyrHisAspAlaAspSerAspLys....................... 302

    654 TCATATTGCAAGTGCGTATTCTTGGCTCGTTGGTGGCAATACCTTTGCAC 703
                 :::||||||:::      |||   :::
    303 .......LysThrAlaTyrThrGlnAlaValThrAsnAlaGluAsnIleL 317

    704 AAAATGGATCAGGTGGTGGCACAGTCAACTTAGGTAGTGAAAAAAATTAAA 753
          |||  ::::::|||:::      ||||||   :::     ::::::
    317 euAsnLysAsnSerGlySer......AsnLeuAspLysThrAlaValGlu 331

    754 CATAGCCCATATGGTTTTTACCAACAGGAGGCTCATTTGGCGACAGTGG 803

    331 ................................................. 331

    804 CTCACCAATGTTTATCTATGATGCCCAAAAGCAAAAGTGGTTAATTAATG 853
                                                   |||:
    332 ...............................................AsnA 333

    854 GGGTATTGCAAACGGGCAACCCCTATATAGGAAAAAGCAATGGCTTCCAG 903
        :::::   |||::::::|||       |||      ||||||
    333 laLeuSerGlnValAlaAsnAlaLys...GlyAlaLeuAsnGly...... 346

    904 CTGGTTCGTAAAGATTGGTTCTATGATGAAATCTTTGCTGGAGATACCCA 953
                                                    :::||
    347 ................................................AsnHi 348

    954 TTCAGTATTCTACGAACCACGTCAAAATGGGAAATACTCTTTTAACGACG 1003
        |::::::               :::::::::|||   :::   |||
    348 sAsnLeu...............GluGlnAlaLysSerAsnAlaAsnThrT 360

   1004 ATAATAATGGCACAGGAAAAATCAATGCCAAACATGAACACAATTCTCTG 1053
             ||||||               :::|||      ::::::
    360 hrIleAsnGly..............LeuGlnHisLeuThrThrAlaGln 371

   1054 CCTAATAGATTAAAAACACGAACCGTTCAATTGTTTAATGTTTCTTTATC 1103
            ::::::||||||   ::::::   |||     ||||||:::
    372 LysAspLysLeuLysGlnGlnValGlnGlnAlaGlnAsnValAlaGlyVa 388

   1104 CGAGACAGCAAGAGAACCTGTTTATCATGCTGCAGGTGGTGTCAACAGTT 1153
          :::|||:::::::::   :::               |||::::::::::::
    388 lAspThrValLysSerSerAlaAsnThrLeuAsnGlyAlaMetGlyThrL 405

   1154 ATCGACCCAGACTG..................AATAATGGAGAAAATATT 1185
          |||       :::                  :::||||||:::|||
    405 euArgAsnSerIleGlnAspAsnThrAlaThrLysAsnGlyGlnAsnTyr 421

   1186 ...................................TCCTTTATTGACGA 1199
                                          :::   :::|||::
    422 LeuAspAlaThrGluArgAsnLysThrAsnTyrAsnAsnAlaValAspSe 438

   1200 AGGAAAAGGCGAATTGATACTTACCAGCAACATCAATCAAGGTGCTGGAG 1249
        ::::::|||   :::        |||||||||   |||:::    |||::::
    438 rAlaAsnGlyValIleAsnAlaThrSerAsnProAsnMetAspAlaAsnA 455

   1250 GATTATATTTCCAAGGAGATTTTACGGTCTCGCCTGAAAATAACGAAACT 1299
        :::::         :::        :::::::|||   :::|||       :::
```

```
 455 laIleAsnGlnIleAlaThrGlnValThrSerThrLysAsn......Ala 469

1300 TGGCAAGGCGCGGGCGTTCATATCAGTGAAGACAGTACCGTTACTTGGAA 1349
         :::|||:::                 :::      :::|||:::|||
 470 LeuAspGlyThrHisAsnLeuThrGlnAlaLysGlnThrAlaThrAsnAl 486

1350 AGTAAACGGCGTGGCAAACGACCGCCTGTCCAAAATCGGCAAAGGCACGC 1399
        ::::::|||::::::|||        |||:::|||     |||   :::|
 486 aIleAspGlyAlaThrAsn......LeuAsnLysAlaGlnLysAspAlaL 501

1400 TGCACGTTCAAGCCAAAGGGGAAAACCAAGGCTCGATCAGCGTGGGCGAC 1449
      ||   :::|||:::   :::   ::::::             |||::::::
 501 euLysAlaGlnValThrSerAlaGlnArg............ValAlaAsn 513

1450 GGTACAGTCATTTTGGATCAGCAGGCAGACGATAAAGGCAAAAAACAAGC 1499
         |||   |||                               :::|||::
 514 ValThrSerIle............................GlnGlnTh 520

1500 CTTTAGTGAAATCGGCTTGGTCAGCGGCAGGGGTACGGTGCAACTGAATG 1549
      :   :::|||::::::   :::   |||:::             ::::::
 520 rAlaAsnGluLeuAsnThrAlaMetGlyGlnLeuGlnHisGlyIleAspA 537

1550 CCGATAATCAGTTCAACCCCGACAAACTCTATTTCGGCTTTCGCGGCGGA 1599
        :::|||      :::     :::   |||         :::   :::
 537 spGluAsnAlaThrLysGlnThrGlnLysTyrArgAspAlaGluGlnSer 553

1600 CGTTTGGATTTAAACGGGCATTCGCTTTCGTTCCACCGT...ATTCAAAA 1646
     :::                ::::::::::::     :::   |||   ||
 554 LysLysThrAlaTyrAspGlnAlaValAlaAlaAlaLysAlaIleLeuAs 570

1647 TACCGATGAAGGGGCGATGATTGTCAACCACAATCAAGACAAA....... 1689
     |    :::   |||:::                |||:::||||||
 570 nLysGlnThrGlySer...............AsnSerAspLysAlaAlaV 582

1690 .....GAATCCACCGTTACCATTACAGGCAATAAAGATATTGCTACAACC 1734
          ::::::          :::|||::::::||||||   |||   :::
 582 alAspArgAlaLeuGlnGlnValThrSerThrLysAsp...AlaLeuAsn 597

1735 GGCAATAACAACAGCTTGGATAGCAAAAAAGAAATTGCCTACAACGGTTG 1784
     |||:::   :::        ::::::|||            |||
 598 GlyAspAlaLysLeuAlaGluAlaLysAlaAlaAlaLysGlnAsn..... 612

1785 GTTTGGCGAGAAAGATACGACCAAAACGAACGGGCGGCTCAACCTTGTTT 1834
                                  |||   ||||||   :::
 613 ...........................LeuGlyThrLeuAsnHisIleT 620

1835 ACCAGCCCGCCGCAGAAGACCGCACCCTGCTGCTTTCCGGCGGAACAAAT 1884
        :::            |||             ::::::   :::|||:::
 620 hrAsnAlaGlnArgThrAspLeuGluGlyGlnIleAsnGlnAlaThrThr 636

1885 TTAAACGGC...AACATCACGCAAACAAACGGCAAACTGTTTTTCAGCGG 1931
     ::::::|||   |||   ::::::||||||::::::   :::
 637 ValAspGlyValAsnThrValLysThrAsnAlaAsnThrLeu........ 650

1932 CAGACCAACACCGCACGCCTACAATCATTTAAACGACCATTGGTCGCAAA 1981
                        |||   |||   |||:::         ::::::|
 651 ..........AspGlyAlaMetAsnSerLeuGlnGlySerIleAsnAspL 664

1982 AAGAGGGCATTCCTCGCGGGGAAATCGTGTGGGACAACGACTGGATCAAC 2031
     ||::::::      |||::::::          |||   |||      :::
 664 ysAspAlaThrLeuArgAsnGlnAsnTyrLeuAspAlaAspGluSerLys 680

2032 CGCACATTTAAAGCGGAAAACTTCCAAATTAAAGGCGGACAGGCGGTGGT 2081
     |||         :::|||   :::      :::   ::::::::::::::::
 681 Arg......AsnAlaTyrThrGlnAlaValThrAlaAlaGluGlyIleLe 695

2082 TTCCCGC............AATGTTGCCAAAGTGAAAGGCGATTGGCATT 2119
     :::::::             |||::::::|||:::      |||      |
 695 uAsnLysGlnThrGlyGlyAsnThrSerLysAlaAspValAspAsnAlaL 712

2120 TGAGC......AATCACGCCCAAGCAGTTTTTGGTGTCGCACCGCATCAA 2163
     ||:::       ::::::|||:::|||:::::::::   |||    :::
 712 euAsnAlaValThrArgAlaLysAlaAlaLeuAsnGlyAlaAspAsn... 727

2164 AGCCACACAATCTGTACACGTTCGGACTGGACGGGTCTGACAAATTGTGT 2213
                           |||:::      |||:::   ||||||
 728 ...............LeuArgAsnAlaLysThrSerAlaThrAsn..... 737

2214 CGAAAAAACCATTACCGACGATAAAGTGATTGCTTCATTGACTAAGACCG 2263
             ||||||              :::::::::|||   |||   :
 738 .......ThrIleAspGlyLeuProAsnLeuThrGlnLeuGlnLysAspA 752

2264 ACATCAGCGGCAATGTCGATCTTGCCGATCACGCTCATTTAAATCTCACA 2313
     :::::::   :::|||:::   |||:::              |||::::::
 752 snLeuLysHisGlnValGluGlnAlaGln............AsnValAla 764

2314 GGGCTT.....................GCCACACTCAACGGCAATCTTAG 2342
     |||:::                          |||||||||     :::::
 765 GlyValAsnGlyValLysAspLysGlyAsnThrLeuAsnThrAlaMetGl 781

2343 TGCA..................AATGGCGATACA...............C 2359
     :|||                   |||   :::|||               :
 781 yAlaLeuArgThrSerIleGlnAsnAspAsnThrThrLysThrSerGlnA 798

2360 GTTATACAGTCAGCCACAACGCCACCCAAAACGGCAACCTTAGCCTCGTG 2409
     ::|||      :::   ::::::::::::|||:::      :::   |||
 798 snTyrLeuAspAlaSerAspSerAsnLysAsnAsnTyrAsnThrAlaVal 814

2410 GGCAATGCCCAAGCAACATTTAATCAAGCCACATTAAACGGCAACACATC 2459
     :::||||||::::::::::::|||   ||||||   |||   |||    ::
 815 AsnAsnAlaAsnGlyValIleAsn...AlaThrAsnAsnProAsnMetAs 830

2460 GGCTTCGGGCAATGCTTCATTTAATCTAAGCGACCACGCCGTACAAAACG 2509
     :|||:::
 830 pAlaAsn........................................ 832

2510 GCAGTCTGACGCTTTCCGGCAACGCT...AAGGCAAACGTAAGCCATTCC 2556
             :::::::::|||   |||   ::::::|||::::::   :::
 833 ........AlaIleAsnGlyMetAlaAsnGlnValAsnThrThrLysAla 846

2557 GCACTCAACGGTAATGTCTCCCTAGCCGATAAGGCAGTATTCCATTTTGA 2606
     ||||||||||||         :::|||||||:::
 847 AlaLeuAsnGlyAlaGlnAsnLeuAlaGln..................Al 857

2607 AAGCAGCCGCTTTACCGGACAAATCAGCGGCGGCAAGGATACGGCATTAC 2656
      :::::::::   |||:::   |||::::::::::   |||
 857 aLysThrAsnAlaThrAsnThrIleAsnAsnAlaHisAspLeuAsnGlnL 874

2657 ACTTAAAAGACAGCGAATGGACGCTGCCGTCAGGCACGGAATTAGGCAAT 2706
            ||||||:::                     |||:::::::::|||
 874 ysGlnLysAspAla...............LeuLysThrGlnValAsnAsn 885

2707 TTA.....................AACCTTGACAACGCCACCATTACACT 2735
                             |||:::::::::::::::      ||
 886 AlaGlnArgValSerAspAlaAsnAsnValGlnHisThrAlaThrGluLe 902

2736 CAATTCCGCCTAT................................CGCC 2752
     ||||||||||                                    |||
 902 uAsnSerAlaMetThrAlaLeuLysAlaAlaIleAlaAspLysGluArgT 919

2753 ACGATGCGGCAGGGGCGCAAACCGGCAGTGCGACAGATGCGCCGCGCCGC 2802
            |||:::|||     :::::::::        |||      :::|||
 919 hrLysAlaSerGlyAsnTyrValAsnAla......AspGlnGluLysArg 933

2803 CGTTCGCGCCGTTCGCGC...........CGTTCCCTATTATCCGTTAC 2840
     ::::::          |||:::         ::::::::::::|||    ||
 934 GlnAlaTyrAspSerLysValThrAsnAlaGluAsnIleIleSerGlyTh 950

2841 ACCGCCAACTTCGGTAGAATCCCGTTTCAACACGCTGACGGTAAACGGC. 2889
     ||||     :::                       |||||||||||||||
 950 rProAsnAla.....................ThrLeuThrValAsnAspV 960

2890 ...................................AAATTGAACGGT 2901
                                          |||||||||
 960 alAsnSerAlaAlaSerGlnValAsnAlaAlaLysThrAlaLeuAsnGly 976
```

```
2902 CAGGGAACATTCCGCTTTATGTCGGAACTCTTCGGCTACCGCAGCGAC.. 2949
     ::::::::::::|||      :::|||     :::        |||
 977 AspAsnAsnLeuArgValAlaLysGluHisAlaAsnAsnThrIleAspGl 993

2950 .........................AAATTGAAGCTGGCGGAAAGTTCCG 2974
                              |||  ||||||  |||:::   :
 993 yLeuAlaGlnLeuAsnAsnAlaGlnLysAlaLysLeuLysGluGlnValG 1010

2975 AAGGCACTTACACCTTG...........GCGGTCAACAATACC...... 3006
     ::::::::   ||||||          :::|||:::|||:::
1010 lnSerAlaThrThrLeuAspGlyValGlnThrValLysAsnSerSerGln 1026

3007 .....................GGCAACGAACCTGCAAGCCTCGAACAATT 3035
                          |||   :::   :::      ::::::
1027 ThrLeuAsnThrAlaMetLysGlyLeuArgAspSerIleAlaAsnGluAl 1043

3036 GACGGTAGTGGAAGGAAAAGACAACAAACCGCTGTCCGAAAACCTTAATT 3085
        |||:::     |||::::::                |||   |||
1043 aThrIleLysAlaGlyGlnAsnTyrThrAspAlaSerProAsn....... 1057

3086 TCACCCTGCAAAACGAACACGTCGATGCCGGCGCGTGGCGTTACCAACTC 3135
             :::||||||:::   :::|||    :::        :::
1058 .....AsnArgAsnGluTyrAspSerAlaValThrAlaAlaLysAlaIle 1072

3136 ATCCGCAAAGACGGCGAGTTCCGCCTGCATAATCCG.............. 3171
     |||::::::    :::                 ||||||
1073 IleAsnGlnThrSer..............AsnProThrMetGluProAs 1084

3172 ..............................GTCAAAGAACAAGAGCTTT 3190
                                   :::|||||||||   |||:
1084 nThrIleThrGlnValThrSerGlnValThrThrLysGluGlnAlaLeuA 1101

3191 CCGACAAACTCGGCAAGGCAGAAGCCAAAAAACAGGCGGAAAAAGACAAC 3240
     ::         :::   |||:::||||||        |||    |||:::|||
1101 snGlyAlaArgAsnLeuAlaGlnAlaLysThrThrAla...LysAsnAsn 1116

3241 GCGCAAAGCCTTGACGCGCTGATTGCGGCCGGGCGCGATGCCGTCGAAAA 3290
        :::::::|||   ::::::         |||   :::||||||:::
1117 LeuAsnAsnLeuThrSerIleAsnAsnAlaGlnLysAspAlaLeu..... 1131

3291 GACAGAAAGCGTT..................................... 3303
        |||:::|||:::
1132 .ThrArgSerIleAspGlyAlaThrThrValAlaGlyValAsnGlnGluT 1148

3303 .................................................. 3303

1148 hrAlaLysAlaThrGluLeuAsnAsnAlaMetHisSerLeuGlnAsnGly 1164

3304 .............................................GCCGA 3308
                                                  |||||
1165 IleAsnAspGluThrGlnThrLysGlnThrGlnLysTyrLeuAspAlaGl 1181

3309 ACCGGCCCGG........................................ 3318
     ||||::::::
1181 uProSerLysLysSerAlaTyrAspGlnAlaValAsnAlaAlaLysAlaI 1198

3319 ........CAGGCAGGCGGGGAAAATGTCGGCATTATGCAGGCGGAGGAA 3360
             :::|||:::|||:::||||||          :::|||:::
1198 leLeuThrLysAlaSerGlyGlnAsnValAspLysAlaAlaValGluGln 1214

3361 GAGAAAAAACGGGTGCAGGCGGATAAAGACACCGCCTTG.........GC 3401
           ::::::|||::::::   |||   |||||||||         ||
1215 AlaLeuGlnAsnValAsnSerThrLys...ThrAlaLeuAsnGlyAspAl 1230

3402 GAAACAGCGCGAAGCG..............GAAACCCGGCCGGCTACCA 3436
     ||||    :::||||||            :::|||       :::    |
1230 aLysLeuAsnGluAlaLysAlaAlaAlaLysGlnThrLeuGlyThrLeuT 1247

3437 CCGCCTTCCCCCGCGCCCGCCGCGCCCGCCGGGATTTGCCGCAACTGCAA 3486
      ||    :::    :::|||:::|||:::         |||
1247 hrHisIleAsnAsnAlaGlnArgThrAlaLeuAsp............... 1258

3487 CCCCAACCGCAGCCCCAACCGCAGCGCGACCTGATCAGCCGTTATGCCAA 3536
```

```
                                  :::   |||::::::   :::||
1259 ...........................AsnGluIleThrGlnAlaThrAs 1266

3537 TAGCGGTTTGAGTGAATTTCCGCCACGCTCAACAGCGTTTTCGCCGTAC 3586
     |                                :::|||:::|||   |||
1266 n.................ValGluGlyValAsnThrValLysAlaLysA 1277

3587 AGGACGAATTAGACCGCGTATTTGCCGAAGACCGCCGCAACGCCGTTTGG 3636
        ::::::||||||   ::::::::::::
1277 laGlnGlnLeuAspGlyAlaMetGlyGln.................Leu 1287

3637 ACAAGCGGCATCCGGGAC..................ACCAAACACTACCG 3668
        ::::::|||||||||                 :::::::::|||::
1288 GluThrSerIleArgAspLysAspThrThrLeuGlnSerGlnAsnTyrGl 1304

3669 TTCGCAAGAT.........TTCCGCGCCTACCGCCAACAAACCGACCTGC 3709
     ::::   |||                  ||||||   |||   ::::::  :::
1304 nAspAlaAspAspAlaLysArgThrAlaTyrSerGlnAlaValAsn.Ala 1320

3710 GCCAAATCGGTATGCAGAAAAACCTCGGCAGCGGGCGCG...TCGGCATC 3756
     |||    ::::::    :::||||||:::::::::::::::::    :::|||::
1321 AlaAlaThrIleLeuAsnLysThrAlaGlyGlyAsnThrProLysAlaAs 1337

3757 CTGTTTTCGCACAACCGGACCGAAAACACCTTCGACGACGGCATCGGCAA 3806
     :         |||:::                  |||    ::::::||||||
1337 pValGluArgAlaMetGlnAlaValThrGlnAlaAsnThrAlaLeuAsnG 1354

3807 CTCGGCACGGCTTGCCCACGGCGCCGTTTTCGGGCAATACGGCATCGACA 3856
          ::::::|||        |||       ::::::|||||||||   |||
1354 lyIleGlnAsnLeuAspArgAlaLysGlnAlaAlaAsnThrAlaIleThr 1370

3857 GGTTCTACATCGGCATCAGCGCGGGCGCGGGTTTAGCAGCGGCAGCCTT 3906
     ::::::::::::    :::::::::     :::                |||
1371 AsnAlaSerAspLeuAsnThrLysGlnLysGluAlaLeuLysAlaGlnVa 1387

3907 TCAGACGGCATCGGAGGCAAAATCCGCCGCCGCGTGCTGCATTACGGCAT 3956
        :::|||:::::::::::::::|||     :::              ||||||
1387 lThrSerAlaGlyArgValSerAlaAlaAsnGlyValGluHisThrAlaT 1404

3957 TCAGGCACGATACCGCGCCGGTTTCG...............GCGGATTCG 3991
        :::    :::||||||    ::::::                ||||||:::
1404 hrGluLeuAsnThrAlaMetThrAlaLeuLysArgAlaIleAlaAspLys 1420

3992 GCATCGAACCGCACATCG.....................GCGCAACGCGC 4020
     |||:::::::::::::|||                    |||::::::
1421 AlaGluThrLysAlaSerGlyAsnTyrValAsnAlaAspAlaAsnLysAr 1437

4021 TATTTCGTCCAAAAAGCGGATTACCGCTACGAAAACGTCAATATCG.... 4066
        :::         :::::::::||||||:::::::::      :::|||
1437 gGlnAlaTyrAspGluLysValThrAlaAlaGluAsnIleValSerGlyT 1454

4067 ..CCACCCCCGGCCTTGCATTCAACCGCTACCGCGCGGGCATTAAGGCAG 4114
        |||    |||:::|||       :::    :::::::::|||       |||
1454 hrProThrProThrLeuThrProAlaAspValThrAsnAlaAlaThrGln 1470

4115 ATTATTCATTCA 4126
      :::   ::::::
1471 ValThrAsnAla 1474
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:AAU37120

```
seq_documentation_block:
ID   AAU37120 standard; Protein; 2344 AA.
XX
AC   AAU37120;
XX
DT   14-FEB-2002  (first entry)
XX
DE   Staphylococcus aureus cellular proliferation protein #1290.
XX
KW   Antisense; prokaryotic cellular proliferation protein;
KW   antibiotic; antibacterial; drug design.
XX
```

```
OS   Staphylococcus aureus.
XX
PN   WO200170955-A2.
XX
PD   27-SEP-2001.
XX
PF   21-MAR-2001; 2001WO-US09180.
XX
PR   21-MAR-2000; 2000US-191078P.
PR   23-MAY-2000; 2000US-206848P.
PR   26-MAY-2000; 2000US-207727P.
PR   23-OCT-2000; 2000US-242578P.
PR   27-NOV-2000; 2000US-253625P.
PR   22-DEC-2000; 2000US-257931P.
PR   16-FEB-2001; 2001US-269308P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Haselbeck R,  Ohlsen KL,  Zyskind JW,  Wall D,  Trawick JD,  Carr GJ;
PI   Yamamoto RT,  Xu HH;
XX
DR   WPI; 2001-611495/70.
DR   N-PSDB; AAS54979.
XX
PT   New polynucleotides for the identification and development of
PT   antibiotics, comprise sequences of antisense nucleic acids -
XX
PS   Example 3; Seq ID No 12713; 511pp; English.
XX
CC   The invention relates to antisense inhibitors of genes essential to
CC   prokaryotic cellular proliferation, their use in identifying the
CC   genes, their use in the discovery of novel antibiotics, the essential
CC   genes themselves and the encoded proteins. The prokaryotes used are
CC   Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC   pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC   invention is also useful for the identification of potential new targets
CC   for antibiotic development. The antisense nucleic acids can also be used
CC   to identify proteins used in proliferation, to express these proteins,
CC   and to obtain antibodies capable of binding to the expressed proteins.
CC   The proteins can be used to screen compounds in rational drug discovery
CC   programmes. The antisense nucleic acid sequence is also useful to screen
CC   for homologous nucleic acids which are required for cell proliferation in
CC   a wide variety of organisms. The present sequence represents an
CC   essential prokaryotic cellular proliferation protein.
CC   Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic
CC   format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   2344 AA;


alignment_scores:
          Quality:   261.50          Length:    1559
            Ratio:   0.361             Gaps:      58
 Percent Similarity:  46.504   Percent Identity: 17.704

alignment_block:
 US-09-303-518D-649 x AAU37120   ..

 Align seg 1/1  to: AAU37120  from: 1  to: 2344

     163 CGCGACTTTGCCGAAAATAAAGGCAAGTTTGCAGTCGGGGCGAAAGATAT 212
           |||   :::   ::::::|||:::                |||   |||:::||
      20 ArgLeuTyrLysSerGlyLysAsnTrpValLysSerGlyIleLysGluIl 36

     213 TGAGGTTTACAACAAAAAAGGGGAGTTGGTCGGCAAATCAATGACAAAAG 262
           ||||:::::::::        |||
      36 eGluMetPheLysIleMetGly..........................L 44

     263 CCCCGATGATTGATTTTTCTGTGGTGTCGCGTAACGGC............ 300
            |||:::|||:::   |||:::||||||::::::::
      44 euProPheIleSerHisSerMetValSerGlnAspAsnGlnSerIleSer 60

     301 ..........................GTGGCGGCATTGGTGGGCGATCA 323
                                   ::::::|||::::::|||
      61 LysLysMetThrGlyTyrGlyLeuLysThrThrAlaValIleGlyGlyAl 77

     324 ATATATTGTGAGCGTGGCACATAAC......................... 348
          :::    |||:::::   |||:::
      77 aPheThrValAsnMetLeuHisAspGlnGlnAlaPheAlaAlaSerAspA 94

     349 ......................................GGCGGCTAT 357
                                                |||:::
      94 laProLeuThrSerGluLeuAsnThrGlnSerGluThrValGlyAsnGln 110

     358 AACAACGTTGATTTTGGTGCGGAAGGAAGAAATCCCGAT........... 396
         |||:::::   :::   |||:::      :::   |||
     111 AsnSerThrThrIleAspAlaSerThrSerThrAlaAspSerThrSerVa 127

     397 ...........................CAACATCGTTTTACTTATAAAA 418
                                     ::::::      |||    :::
     127 lThrLysAsnSerSerSerValGlnThrSerAsnSerAspThrValSerS 144

     419 TTGTGAAACGGAATAATTATAAAGCAGGGACTAAAGGCCATCCTTATGGC 468
                 |||   :::|||       :::   |||::::::
     144 erGluLysSerGluAsnValThrSerThrThrAsnSer............ 156

     469 GGCGATTATCATATGCCGCGTTTGCATAAATTTGTCACAGATGCAGAACC 518
                                              |||:::    :::
     157 .................................ThrSerAsnGlnGl 161

     519 TGTTGAAATGACCAGTTATATGGATGGGCGGAAATATATCGATCAAAATA 568
            ::::::|||||     ::::::   :::       :::   ::::
     161 nGluLysLeuThrSerThrSerGluSerThrSerSerLysAsnThrThrS 178

     569 ATTACCCTGACCGTGTTCGTATTGGGGCAGGCAGGCAATATTGGCGATCT 618
         ::      |||      :::   :::                  |||
     178 erSerSerAspThrLysSerValThrSer..............ThrSer 189

     619 GATGAAGATGAGCCCAATAACCGCGAAAGTTCATATCATATTGCAAGT.. 666
         :::   ::::::|||   |||   :::::::::         ||||||
     190 SerThrGluGlnProIleAsnThrSerThrAsnGlnSerThrAlaSerAs 206

     666 .................................................. 666

     206 nAsnThrSerGlnSerThrThrProThrSerAlaAsnLeuAsnLysThrS 223

     667 ...................................GCGTATTCTTGG 678
                                             ::::::|||
     223 erThrThrSerThrSerThrAlaProValLysLeuArgThrPheSerArg 239

     679 CTCGTTGGTGGCAATACCTTTGCACAAAATGGATCA...........GG 716
         |||    :::    :::||||||||||:::     ::::::
     240 Leu...AlaMetSerThrPheAlaSerAlaAlaThrThrThrAlaLeuTh 255

     717 TGGTGGCACAGTCAACTTAGGTAGTGAAAAAATTAAACATAGCCCATATG 766
           ::::::|||:::::::::::::::::::::::|||:::
     255 rAlaAsnThrIleThrValAsnLysAspAsnLeuLysGln.......... 268

     767 GTTTTTACCAACAGGAGGCTCATTT.........GGCGACAGTGGCTCA 807
           ::::::   |||:::|||:::                 ::::::|||
     269 ..TyrMetThrThrSerGlyAsnAlaThrTyrAspGlnSerThrGlyVal 284

     808 CCAATGTTTATCTATGATGCCCAAAAGCAAAAGTGGTTAATT........ 849
             :::          |||:::    :::||||||      |||
     285 ValThrLeuThrGlnAspThrThrSerGlnLysGlyAlaIleThrLeuGl 301

     850 ..................................AATGGGGTATTGCAAA 865
                                            :::|||   ::::::
     301 yThrArgIleAspSerAsnLysSerPheHisPheSerGlyLysValAsnL 318

     866 CGGGCAACCCCTATATAGGAAAAAGCAATGGCTTCCAGCTGGTTCGTAAA 915
           ||||||    |||    |||    :::||||||
     318 euGlyAsnLysTyrGluGlyHisGlyAsnGly.................. 328
```

```
 916 GATTGGTTCTATGATGAAATCTTTGCTGGAGATACCCATTCAGTATTCTA 965
                                 ||||||    :::    ::
 329 ..........................GlyAspGlyIleGlyPheAlaPh 336

 966 CGAACCACGTCAAAATGGGAAATACTCTTTTAACGACGATAATAATGGCA 1015
     :::|||           |||:::  ::::::|||          |||
 336 eSerProGlyValLeuGlyGluThrGlyLeuAsnGlyAlaAlaValGlyI 353

1016 CAGGA........................AAAATCAATGCCAAACATGAA 1041
        |||                       |||:::::::::   |||
 353 leGlyGlyLeuSerAsnAlaPheGlyPheLysLeuAspThrTyrHis... 368

1042 CACAATTCTCTGCCTAATAGATTAAAAACACGAACCGTTCAATTGTTTAA 1091
     ::::|||   ||||||
 369 AsnThrSerThrProAsn................................ 374

1092 TGTTTCTTTATCCGAGACAGCAAGAGAACCTGTTTATCATGCTGCAGGTG 1141
         |||    :::::::::      :::|||           |||:::||||
 375 ....SerSerAlaLysAlaLysAlaAspProSerAsnValAlaGlyGlyG 390

1142 GTGTCAACAGTTATCGACCCAGACTGAATAATGGAGAAAATATTTCCTTT 1191
     ||:::                            |||     ::::::
 390 lyAla.......................PheGlyAlaPheValThrThr 398

1192 ATTGACGAAGGAAAAGGCGAATTGATACTTACCAGCAACATCAATCAAGG 1241
        :::    |||   ::::::          :::|||:::      :::::
 399 AspSerTyrGlyValAlaSerThrTyrThrSerSerSerThrAlaAspAs 415

1242 TGCTGGAGGATTATATTTCCAAGGAGATTTTACGGTCTCGCCTGAAAATA 1291
     :|||:::    |||                    :::|||:::|||   ||||
 415 nAlaAlaLysLeu..................AsnValGlnProThrAsnA 426

1292 ACGAAACTTGGCAAGGCGCGGGCGTTCATATCAGTGAAGACAGTACCGTT 1341
     ||   |||:::|||          ::::::   :::   |||:::    |||
 426 sn...ThrPheGlnAspPheAspIleAsnTyrAsnGlyAspThrLysVal 441

1342 .......................ACTTGGAAAGTAAACGGCGTGGCAAA 1367
                            ||||||      |||          ::
 442 MetThrValLysTyrAlaGlyGlnThrTrpThrArgAsn......IleSe 456

1368 CGACCGCCTGTCCAAAATCGGCAAAGGCACGCTGCACGTTCAAGCCAAAG 1417
     :|||    ::::::|||   |||      ::::::   ::::::       :
 456 rAspTrpIleAlaLysSerGlyThrThrAsnPheSerLeuSerMetThrA 473

1418 GGGAAAACCAAGGCTCGATCAGC.........GTGGGCGACGGTACAGTC 1458
     ::::::::    |||:::   :::              |||       ||||||
 473 laSerThrGlyGlyAlaThrAsnLeuGlnGlnValGlnPheGlyThrPhe 489

1459 ATTTGGATCAGCAGGCAGACGATAAAGGCAAAAAACAAGCCTTTAGTGA 1508
             ::::::|||
 490 GluTyrThrGluSerAla............................... 495

1509 AATCGGCTTGGTCAGCGGCAGGGGTACGGTGCAACTGAATGCCGATAATC 1558
                |||:::         |||::::::::::::   ::::
 496 ..........ValThrGlnValArgTyrValAspValThrThrGlyLysA 509

1559 AGTTCAACCCCGACAAACTCTATTTCGGCTTTCGCGGCGGACGTTTGGAT 1608
     :::::   |||   |||    |||    |||
 509 spIleIleProProLysThrTyrSerGly..................... 518

1609 TTAAACGGGCATTCGCTTTCGTTCCACCGTATTCAAAATACCGATGAAGG 1658
                                              |||:::|||:::
 519 .................................AsnValAspGlnVa 523

1659 GGCGATGATTGTCAACCACAATCAAGACAAAGAATCCACCGTTACCATTA 1708
       :::   |||   |||              ::::::|||::::::|||
 523 lValThrIleAspAsn............GlnGlnSerAlaLeuThrAlaL 536

1709 CAGGCAATAAAGATATTGCTACAACCGGCAATAACAACAGCTTGGATAGC 1758
        |||                           |||   :::|||:::||||||
 536 ysGlyTyr.....................AsnTyrThrSerValAspSer 545

1759 AAAAAAGAAATTGCCTACAACGGTTGGTTTGGCGAGAAAGATACGACCAA 1808
```

```
     :::            :::||||||                    ||||||:::||
 546 SerTyrAlaSerThrTyrAsn.................AspThrAsnLy 556

1809 A...........ACGAACGGGCGGCTCAACCTTGTTTAC.......... 1836
     |           |||||||:::      ::::::::::|||
 556 sThrValLysMetThrAsnAlaGlyGlnSerValThrTyrTyrPheThrA 573

1837 ........CAGCCCGCCGCAGAAGACCGCACCCTGCTGCTTTCCGGCGGA 1878
                |||::::::           :::     ::::::    |||
 573 spValLysAlaProThrValThrValGlyAsnGlnThrIleGluValGly 589

1879 ACAAATTTAAACGGCAACATC......ACGCAAACAAACGGCAAACTGTT 1922
         ::::::|||     :::        |||      ||||||
 590 LysThrMetAsnProIleValLeuThrThrThrAspAsnGly........ 603

1923 TTTCAGCGGCAGACCAACACCGCACGCCTACAATCATTTAAACGACCATT 1972
         :::|||     |||
 604 ....ThrGlyThrValThr............................... 608

1973 GGTCGCAAAAAGAGGGCATTCCTCGCGGGGAAATCGTGTGGGACAACGAC 2022
       :::               |||:::|||    |||
 609 ..AsnThrValThrGlyLeuProSerGly..................... 617

2023 TGGATCAACCGCACATTTAAAGCGGAAAACTTCCAAATTAAAGGCGGACA 2072
              ::::::    :::    :::    :::|||    |||
 618 .........LeuSerTyrAspSerAlaThrAsnSerIleIleGlyThrPr 631

2073 GGCGGTGGTTTCCCGCAATGTTGCCAAAGTGAAAGGCGATTGGCATTTGA 2122
      :::    ::::::::::::::::::    |||    :::        :::    :
 631 oThrLysIleGlyGlnSerThrValThrValValSerThrAspGlnAlaA 648

2123 GCAATCACGCCCAAGCAGTTTTTGGTGTC.................... 2151
     ::|||    :::    ::::::|||    :::
 648 snAsnLysSerThrThrThrPheThrIleAsnValValAspThrThrAla 664

2152 .........................................GCACCGCA 2159
                                               :::|||
 665 ProThrValThrProIleGlyAspLysSerSerGluValPheSerProIl 681

2160 TCAAAGCCACACAATCTGTACACGTTCGGACTGGACGGGTCTGACAAATT 2209
      :::        :::|||:::|||:::    |||    :::|||       |||:
 681 eSerProIleAsnIleAlaThrGln...AspAsnSerGly......AsnA 695

2210 GTGTCGAAAAAACCATTACCGACGATAAAGTGATTGCTTCATTGACTAAG 2259
     ::|||    :::|||:::|||                 :::    |||
 695 laValThrAsnThrValThrGly.........LeuProSerGlyLeuThr 708

2260 ACCGACATCAGCGGCAATGTCGATCTTGCCGATCACGCTCATTTAAATCT 2309
        |||    ::::::|||:::         :::       ::::::::::::
 709 PheAspSerThrAsnAsnThrIleSerGlyThrProThrAsnIleGlyTh 725

2310 CACAGGGCTTGCCACACTCAACGGCAATCTTAGTGCAAATGGCGATACA. 2358
      :::    ::::::    ::::::    :::    |||:::|||       |||
 725 rSerThrIleThrIleValSerThrAspAlaSerGlyAsnLysThrThrT 742

2359 ........CGTTATACAGTCAGCCACAACGCCACCCAAAACGGCAACCTT 2400
             :::|||    |||::::::|||    :::::::::    ::::::
 742 hrThrPheLysTyrGluValThrArgAsn...SerMetSerAspSerVal 757

2401 AGCCTCGTGGGCAATGCCCAA...............GCAACATTTAATCA 2435
      |||     |||:::::::|||                  :::|||    :::
 758 SerThrSerGlySerThrGlnGlnSerGlnSerValSerThrSerLysAl 774

2436 AGCCACATTAAACGGCAACACATCGGCTTCGGGCAATGCTTCATTTAATC 2485
        :::    :::::::::||||||:::||||||:::         :::
 774 aAspSerGlnSerAlaSerThrSerThrSerGlySerIleMetThrSerT 791

2486 TAAGCGACCACGCCGTACAAAACGGCAGTCTGACGCTTTCCGGCAACGCT 2535
      |||       :::    ::::::    |||::::::||||||    ::::::
 791 hrSerAlaSerThrSerLysSerThrSerValSerLeuSerAspSerVal 807

2536 AAGGCAAACGTAAGCCATTCCGCACTCAACGGTAATGTCTCCCTAGCCGA 2585
     :::|||:::    |||    |||:::      ::::::|||
```

```
 808 SerAlaSerLysSerLeuSerThrSerGluSerAsn.............. 819

2586 TAAGGCAGTATTCCATTTTGAAAGCAGCCGCTTTACCGGACAAATCAGCG 2635
                                             :::::::|||:
 820 ......................................SerValSerS 823

2636 GCGGCAAGGATACGGCATTACACTTAAAAGACAGCGAATGGACGCTGCCG 2685
     :::::   :::|||:::|||      :::::::|||     :::
 823 erSerThrSerThrSerLeuValAsnSerGlnSerValSerSerSerMet 839

2686 TCAGGCACGGAATTAGGCAATTTAAACCTTGACAACGCCACCATTACACT 2735
     ||||||:::        :::   :::|||::::::          :::
 840 SerGlySerValSerLysSerThrSerLeuSerAspPheIleSerAsnSe 856

2736 CAATTCCGCCTATCGCCACGATGCGGCAGGGGCGCAAACCGGC....... 2778
       :::|||:::   :::   :::::::::::::::::|||:::
 856 rSerSerThrGluLysSerGluSerValSerThrSerThrSerAspSerL 873

2779 ..............AGTGCGACAGATGCGCCGCGCCGCCGTTCGCGCCGT 2814
                   |||   :::|||:::             :::
 873 euArgThrSerThrSerLeuSerAspSerValSerMetSerThrSerGly 889

2815 TCGCGCCGTTCCCTATTATCCGTTACACCGCCAACTTCGGTAGAATCCCG 2864
     |||       :::      |||::::::     ||||||   ::::::
 890 SerLeuSerLysSerGlnSerLeuSerThrSerThrSerAspSerAlaSe 906

2865 TTTCAACACGCTGACGGTAAACGGCAAATTGAACGGTCAGGGAACATTCC 2914
           :::      :::   ::::::      |||:::   :::|||   :
 906 rThrSerGlnSerValSerAspSerThrSerAsnSerIleSerThrSerG 923

2915 GCTTTATGTCGGAACTCTTCGGCTACCGCAGCGACAAATTGAAGCTGGCG 2964
     ::   :::||||||         |||        |||::::::::::::::::::
 923 luSerLeuSerGlu...SerGlySerThrSerGluSerIleSerIleSer 938

2965 GAAAGTTCCGAAGGCACTTACACCTTGGCGGTCAACAATACCGGCAACGA 3014
     :::|||   :::::::::   :::   ::::::::::::      :::::
 939 AsnSerIleSerAsnSerValSerAlaSerThrSerLysLeuGluSerGl 955

3015 ACCTGCAAGCCTCGAACAATTGACGGTAGTGGAAGGAAAAGACAACAAAC 3064
     :   :::|||::::::        |||               :::|||:::|||
 955 nSerThrSerIleSerLeuSerThr............SerAspSerLysS 968

3065 CGCTG......TCCGAAAACCTTAATTTCACCCTGCAAAACGAACACGTC 3108
        :::         ||||||:::|||:::    :::     :::::::::
 968 erMetSerThrSerGluSerLeuSerAspSerThrSerThrSerAspSer 984

3109 GATGCCGGCGCGTGGCGTTACCAACTCATCCGCAAAGACGGCGAGTTCCG 3158
        :::|||:::
 985 ValSerGlySerLeu................................. 989

3159 CCTGCATAATCCGGTCAAAGAACAAGAGCTTTCCGACAAACTCGGCAAGG 3208
                :::    :::    :::|||::::::|||    :::    :::    :
 990 .......SerValAlaGlySerGlnSerValSerThrSerThrSerAspS 1004

3209 CAGAAGCCAAAAAACAGGCGGAAAAAGACAACGCGCAAAGCCTTGACGCG 3258
     ::    :::    ::::::       :::|||:::    ::::::           :::
1004 erMetSerThrSerGluMetIleSerAspSerMetSerThrSerGlySer 1020

3259 CTGATTGCGGCCGGGCGCGATGCCGTCGAAAAGACAGAAAGCGTTGCCGA 3308
     |||    |||:::          :::::::::    ::::::|||::::::
1021 LeuAlaAlaSerAspSerLysSerMetSerValSerSerSerMetSerTh 1037

3309 ACCGGCCCGGCAGGCAGGCGGGGAAAATGTC......GGCATTATGCAGG 3352
                  :::    :::|||::::::      :::|||          :
1037 rSerGlnSerGlySerThrSerGluSerLeuSerAspSerIleSerThrS 1054

3353 CGGAGGAAGAGAAAAAACGGGTGCAGGCGGATAAAGACACCGCCTTGGCG 3402
     ::::::::::::|||    :::                  :::|||:::
1054 erAspSerAspSerLysSerLeu.................SerLeuSer 1064

3403 AAACAGCGCGAAGCGGAAACCCGGCCGGCTACCACCGCCTTCCCCCGCGC 3452
        :::::::::::::|||         :::|||::::::          ::
1065 ThrSerGlnSerGlySerThrSerThrSerThrSerThrSerSerSerVa 1081

3453 CCGCCGCGCCCGCCGGGATTTGCCGCAACTGCAACCCCAACCGCAGCCCC 3502
     :|||
1081 lArg.............................................. 1082

3503 AACCGCAGCGCGACCTGATCAGCCGTTATGCCAATAGCGGTTTGAGTGAA 3552
                                              :::||||||
1083 .........................................MetSerGlu 1085

3553 TTTTCCGCCACGCTCAACAGCGTTTTCGCCGTACAGGACGAATTAGACCG 3602
        ::::::|||   :::|||:::   :::      |||::::::     |||
1086 SerGlnSerThrSerGlySerMetSerThrSerGlnSerAspSer.Thr. 1101

3603 CGTATTTGCCGAAGACCGCCGCAACGCCGTTTGGACAAGCGGCATCCGGG 3652
                    :::    :::|||    |||::::::::::::|||
1102 ..............SerIleSerThrSerPheSerAspSerThrSerAsp 1113

3653 ACACCAAACACTACCGTTCGCAAGATTTCCGCGCCTACCGCCAACAAACC 3702
     :::                          ||||||   ||||||::::::
1114 SerLys...................SerAlaSerThrAlaSerSerGl 1123

3703 GACCTGCGCCAAATCGGTATGCAGAAAAACCTCGGCAGCGGGCGCGTCGG 3752
       :::   ::::::||||||       :::||||||::::::    ::::::
1123 uSerIleSerGlnSerValSerThrSerThrSerGlySerValSerThrS 1140

3753 CATCCTGTTTTCGCACAACCGGACCGAAAACACCTTCGACGACGGCATCG 3802
     :::::   :::     |||::::::    :::          ||||||:::    |||
1140 erThrSerLeuSerThrSerAsnSerGluArgThrSerThrSerMetSer 1156

3803 GCAACTCGGCACGGCTTGCCCACGGCGCCGTTTTCGGGCAATACGGCATC 3852
        :::        :::|||    |||:::          |||    :::|||::::
1157 AspSerThr...SerLeuSerThrSerGlu...SerAspSerThrSerAs 1171

3853 GACAGGTTCTACATCGGCATCAGCGCGGGCGCGGGTT............ 3889
     ::::    |||||||||    |||                :::
1171 pSerThrSerThrSerAspSerIleSerGluAlaIleSerGlySerGluS 1188

3890 .............TTAGCAGCGGCAGCCTTTCAGACGGCATCGGAGGCA 3925
                  |||:::    :::             :::    ||||||:::
1188 erThrSerIleSerLeuSerGluSerAsnSerThrSerAspSerGluSer 1204

3926 AAATCCGCCGCCGCGTGCTGCATTACGGCATTCAGGCACGATACCGCGCC 3975
     |||||||||:::|||                ::::::    ::::::|||:::
1205 LysSerAlaSerAlaPheLeuSerGluSerLeuSerGluSerThrSerGl 1221

3976 GGTTTCGGCGGATTCGGCATCGAACCGCACATCGGCGCAACGCGCTATTT 4025
        :::::::::|||    |||:::    ||||||    :::    ::::::|
1221 uSerThrSerGluSerLeuSerGlySerThrSerAspSerThrSerLeuS 1238

4026 CG...TCCAAAAAGCGGATTACCGCTACGAAAACGTCAATATCG...... 4066
     ||    |||:::::::::       :::|||:::||||||:::|||
1238 erAspSerAsnSerGluSerGlySerThrSerThrSerLeuSerAsnSer 1254

4067 ...CCACCCCCGGCCTTGCATTCAACCGCTACCGCGCG........... 4101
               |||       ||||||:::|||:::
1255 ThrSerGlySerAlaSerIleSerThrSerThrSerGlySerAlaSerTh 1271

4102 ..................GGCATTAAGGCAGATTATTCATTCAAACCGG 4132
                       :::::::::::::::     |||    :::    :
1271 rSerThrValLysSerGluSerValSerThrSerLeuSerThrSerThrS 1288

4133 CGCAACACATTTCCATCACGCCTTATTTGAGCCTGTCCTATACCGATGCC 4182
     ::        :::|||    :::          ||||||    |||     :::|||:::
1288 erThrSerLeuSerAspSerThrSerLeuSerThrSerLeuSerAspSer 1304

4183 GCTTCGGGCAAAGTCCGAACACGCGTCAATACCGCCGTATTGGCTCAGGA 4232
     :::||||||:::        :::      ::::::::::::::    ::::::||
1305 ThrSerGlySerLysSerAsnSerLeuSerAlaSerMetSerThrSerAs 1321

4233 TTTCGGCAAAACCCGCAGTGCGGAA 4257
     |         :::|||||||||::::::|||
1321 pSerIleSerThrArgLysSerGlu 1329
```

```
seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01843

seq_documentation_block:
ID   AAB01843 standard; Protein; 992 AA.
XX
AC   AAB01843;
XX
DT   11-SEP-2000  (first entry)
XX
DE   Haemophilus influenzae strain 15 mature HMW1A protein, SEQ ID NO:61.
XX
KW   Mature HMW protein; hmw gene; hmwA1; hmwA2; high molecular weight;
KW   non-typeable Haemophilus influenzae; NTHi; non-encapsulated;
KW   recombinant production; Escherichia coli; antibacterial; vaccine;
KW   human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;
KW   detection; diagnosis.
XX
OS   Haemophilus influenzae strain 15.
XX
PN   WO200020609-A2.
XX
PD   13-APR-2000.
XX
PF   07-OCT-1999;   99WO-CA00938.
XX
PR   07-OCT-1998;   98US-0167568.
PR   08-DEC-1998;   98US-0206942.
XX
PA   (CONN-) CONNAUGHT LAB LTD.
XX
PI   Loosmore SM,  Yang Y,  Klein MH;
XX
DR   WPI; 2000-303789/26.
DR   N-PSDB; AAA52192.
XX
PT   Nucleic acid molecule for producing recombinant high molecular weight
PT   proteins of Haemophilus which are used as a vaccine to provide
PT   protection against Haemophilus induced diseases in humans -
XX
PS   Claim 8; Fig 26A-O; 307pp; English.
XX
CC   The invention relates to the recombinant production of Haemophilus
CC   influenzae high molecular weight (HMW) proteins in Escherichia coli. The
CC   expression construct used to effect recombinant expression comprises a
CC   promoter functional in E. coli (e.g., the T7 promoter) operably linked
CC   to a modified hmwABC operon from a non-typeable (non-encapsulated) H.
CC   influenzae (NTHi). Most HMW-expressing NTHi strains contain two hmw gene
CC   clusters termed hmw1ABC and hmw2ABC. Each hmwABC operon comprises hmwA,
CC   hmwB and hmwC genes. The hmwA genes encode the structural HMWA proteins
CC   and the hmwB and hmwC genes encode accessory proteins which are
CC   responsible for post-translational processing and secretion of the HMWA
CC   proteins. The modified hmwABC operon used in the expression construct of
CC   the invention contains an A gene modified such that it encodes only the
CC   mature HMWA. The invention also discloses hmwA genes (AAA52175-A52198)
CC   and HMWA proteins (AAB01824-B01849) from the non-typeable H. influenzae
CC   strains Joyc, K1, K21, LCDC2, PMH1, 15 and 12. The nucleic acids and
CC   vectors are used for the production of recombinant H. influenzae HMW
CC   proteins which can be used as vaccines to mediate a humoral or
CC   cell-mediated immune response to provide protection against diseases in
CC   humans caused by H. influenzae (e.g., otitis media, epiglottitis,
CC   pneumonia and tracheobronchitis). The HMW proteins are also useful as
CC   antigens in immunoassays for detecting antibodies against Haemophilus,
CC   HMW proteins and/or HMW peptides. The nucleotide sequences encoding the
CC   HMW proteins can be used to isolate and clone hmw genes from other
CC   non-typeable strains of Haemophilus via hybridisation reactions. The
CC   present sequence represents a mature HMWA protein from a non-typeable
CC   strain of H. influenzae.
XX
SQ   Sequence   992 AA;

alignment_scores:
          Quality:  261.00          Length:   1160
```

```
            Ratio:  0.476             Gaps:     58
 Percent Similarity:  47.241   Percent Identity: 19.914

alignment_block:
 US-09-303-518D-649 x AAB01843   ..

 Align seg 1/1  to: AAB01843  from: 1  to: 992

   508 GATGCAGAACCTGTTGAAATGACCAGTTATATGGATGGGCGGAAATATAT 557
       |||       :::|||      ||||||     |||:::      ||
     2 AspAsnValThrIleGluAlaProSerTyrSerArgGlyAsnAlaGlyIl 18

   558 CGATCAAAATAATTACCCTGACCGTGTTCGTATTGGGGCAGGC...AGGC 604
       ||||  :::::::::|||                |||:::|||   ::::
    18 eAsp...SerGluPheProGly...........GlySerGlyThrLysG 30

   605 AATATTGGCGATCTGATGAAGATGAGCCC.........AATAACCGCGAA 645
       ::     :::::::::   :::::|||            :::|||:::
    30 luSerProLysThrAsnGlyGluGlnProThrValLeuThrAsnGluThr 46

   646 AGTTCATATCATATTGCAAGTGCGTATTCTTGGCTCGTT........... 684
          |||   ::::::   |||:::   :::|||::::::
    47 IleSerAsnTyrLeuLysSerGly...ThrTrpValMetAsnIleThrAl 62

   684 .................................................. 684

    62 aLysLysAsnLeuThrValAsnSerSerIleAsnIleGlyAspSerSerH 79

   684 .................................................. 684

    79 isLeuIleLeuHisSerGluGlyLysAsnAsnGlyGlyValLysIleLys 95

   685 ..................GGTGGCAATACCTTTGCACAAAATGGATCA.. 714
                         |||||||||        |||:::|||:::
    96 GluAspIleThrSerAsnGlyGlyAsnLeuThrIleGlnSerGlyGlyTr 112

   715 ...........................GGTGGTGGCACAGTCAACTTA. 735
                                 |||   ||||||:::|||:::
   112 pValAspValHisLysAsnIleThrLeuGlyThrGlyThrLeuAsnIleT 129

   736 ................................GGTAGTGAAAAAATTAAA 753
                                     |||:::||||||   :::
   129 hrAlaLysGlySerIleAlaPheGluGlyAsnGlyThrGluLysAlaArg 145

   754 CATAGCCCATATGGTTTTTACCAACAGGAGGCTCATTTGGCGACAGTGG 803
       ::::::    :::   :::   :::   |||::::::   ::::::||
   146 AsnAlaSerSerAlaGlnIleThrAlaGlnGlyThrIleThrAsnThrGl 162

   804 CTCACCAATGTTTATCTATGATGCCCAAAAGCAAAAGTGGTTA....... 846
       |:::                  |||||||||      |||
   162 yAsp....................GlnLysGlnLeuArgLeuAsnAsnV 172

   847 .....ATTAATGGGGTATTGCAAACGGGCAACCCCTATATAGGAAAAAGC 891
            |||||||||        ||||||    .    ||||||   :::
   172 alSerIleAsnGly.........ThrGly.........IleGlyLeuAsn 182

   892 AATGGCTTCCAGCTGGTTCGTAAAGATTGGTTCTATGATGAAATCTTTGC 941

   182 .................................................. 182

   942 TGGAGATACCCATTCAGTATTCTACGAACCACGTCAAAATGGGAAATACT 991
                         |||   :::   :::   |||   ::::::
   183 ..................PheValSerIleGlnProAsnThrSerHisA 193

   992 CTTTTAACGACGATAATAATGGCACAGGAAAAATCAATGCCAAACATGAA 1041
          |||:::   :::         :::|||::::::         |||
   193 rgPheAspGlyGluLeuIleIleSerGlyArgVal.........HisVal 206

  1042 CACAATTCTCTGCCTAATAGATTA.........AAAACACGAACCGTTCA 1082
       :::::::::   |||:::::|||          |||:::        ::
   207 AsnGlnThrThrProLysAsnLeuSerPheTrpLysValSerAspGluSe 223

  1083 ATTGTTTAATGTTTCTTTATCCGAGACAGCAAGAGAACCTGTTTATCATG 1132
```

```
             :   :::|||||||||          |||::::::|||            :
 223 rTyrTrpAsnValSer...HisLeuThrValLysGlu.........LysS 236

1133 CTGCAGGTGGTGTCAACAGTTATCGACCCAGACTGAATAATGGAGAAAAT 1182
     ::|||   :::   ::::::::      :::   |||:::|||::::::
 236 erAlaPheSerPheThrLysPheAlaLeuAsnAsnAsnHisGlyArgGlu 252

1183 ATTTCCTTTATTGACGAAGGAAAAGGCGAATTGATACTTACCAGCAACAT 1232
                                                |||||||:::
 253 ..................................ThrSerArgTy 256

1233 CAATCAAGGTGCTGGAGGATTATATTTCCAAGGA...............G 1267
      :::::::|||   ||||||:::   |||::::::              :
 256 rArgLysGly...GlyGlyValIlePheArgSerProThrGlyHisThrA 272

1268 ATTTTACGGTCTCGCCTGAAAATAACGAAACTTGGCAAGGCGCGGGCGTT 1317
     ::||||||||||:::                       ||||||:::   :::
 272 snPheThrValLys....................GlnGlySerValAla 281

1318 CATATCAGT............GAAGACAGTACCGTTACTTGGAAAGTAAA 1355
     ::::::|||             :::|||::::::   :::   ::::::
 282 AsnPheSerPheLysAlaLysAsnAspThrAsnHisAlaAsnGlnLeuPr 298

1356 CGGCGTGGCAAACGACCGCCTGTCCAAAATCGGCAAAGGCACGCTG..... 1401
                 |||:::::::|||         |||   |||    :::
 298 oIleGlnPheAsnSerAsnIleSerValAspGlyGlyGlyLysValLeuP 315

1402 ...................CACGTTCAAGCCAAAGGGGAAAACCAAGGC 1431
                                 ::::::   |||    :::::::::
 315 heCysIleThrSerAsnTyrSerGlyArgSerValGlyIleGlyMetSer 331

1432 TCGATCAGCGTGGGCGACGGTACAGTCATTTTGGATCAGCAGGCAGACGA 1481
     ||||||:::|||:::||||||:::   :::     :::::::::
 332 SerIleAsnValSerAspGlySerAsnLeuThrPheAsnSerSer..... 346

1482 TAAAGGCAAAAAACAAGCCTTTAGTGAAATCGGCTTGGTCAGCGGCAGGG 1531
                                           :::   |||:::
 347 ..................................IleArgGlyGlnG 351

1532 GTACGGTGCAACTGAATGCCGATAATCAGTTCAACCCCGACAAACTCTAT 1581
       :::   ::::::::   |||        :::|||           :::
 351 luAlaPheAsnIleSerLysAspLeuThrIleAsnAlaThrGlySerPhe 367

1582 TTCGGCTTTCGCGGCGGACGTTTGGATTTAAACGGGCATTCGCTTTCGTT 1631
     |||    :::                 :::||||||:::::::::::::
 368 PheGluLeuGlyGlnTyrSerAspThrPheAsnGlyAsnGlyPheAsnHi 384

1632 CCACCGTATTCAAAATACCGATGAA...........GGGGCGATGATTG 1669
             |||::::::|||    :::            |||:::    ::::
 384 sAspAlaIleLysSerThrHisAsnIleSerIleLeuGlyGlyAsnValT 401

1670 TCAACCACAATCAAGACAAAGAATCCACCGTTACCATTACAGGCAATAAA 1719
     ::         :::|||||||::::::|||        ||||||||||||||
 401 hrLeuGlyGlyGlnAspSerSerSer......ThrIleThrGlyAsnIle 415

1720 GATATTGCTACAACCGGCAATAACAACAGCTTGGATAGCAAAAAAGAAAT 1769
     :::|||:::   ::::::|||    :::
 416 AsnIleSerGlnAlaAlaAsnValThrLeuArg................ 426

1770 TGCCTACAACGGTTGGTTTGGCGAGAAAGATACGACCAAAACGAACGGGC 1819
      |||||||||||               :::|||:::   |||      |||
 427 .AlaTyrAsnGlyAsnGlyArgAsnLysGlnLeuThrLeuGlyAsn.... 441

1820 GGCTCAACCTTGTTTACCAGCCCGCCGCAGAAGACCGCACCCTGCTGCTT 1869
                               :::   |||   :::      |||:::
 442 ...................ValSerIleGluGlyAsnLeuSerLeuIle 451

1870 TCCGGCGGAACAAATTTAAACGGCAACATCACG...CAAACAAACGGCAA 1916
     ::::::::::::|||:::||||||||::::::   :::   |||:::||
 452 GlyAlaSerAlaAsnIleAsnGlyAsnLeuSerValLysGluAsnAlaLy 468

1917 ACTGTTTTTCAGCGGCAGACCA..................ACACCGCACG 1948
     |       |||:::|||:::                    |||

 468 s......PheLysGlyGluThrGlnAspAsnLeuAsnIleThrGlyThrP 483

1949 CCTACAATCATTTAAACGACCATTGGTCGCAAAAAGAGGGCATTCCTCGC 1998
           |||:::   ::::::      :::   ::::::|||:::   :::
 483 heIleAsnAsnGlyAspSerLysIleAsnIleSerGlnGlyValValLys 499

1999 ...GGGGAAATCGTGTGGGACAACGACTGGATCAACCGCACATTTAAAGC 2045
        |||:::::::::   |||:::|||   :::|||   |||      ||
 500 LeuGlyAsnValThrAsnAspGlyAsp...LeuAsnIleThrThrHisAl 515

2046 GGAAAACTTCCAA......ATTAAAGGCGGACAGGCGGTG.......... 2079
     |::::::   |||       |||   ||||||:::   :::
 515 aLysHisAsnGlnArgSerIleIleGlyGlyAspIleIleAsnLysLysG 532

2080 ........................................GTTTCC 2085
                                                :::::
 532 lySerLeuAsnIleThrAspSerAsnLysAsnAlaGluIleGlnIleGly 548

2086 CGCAATGTTGCCAAAGTGAAAGGCGATTGG.................... 2115
         |||:::::::::   :::|||:::
 549 GlyAsnIleSerGlnLysGluGlyAsnLeuThrIleSerSerAspLysIl 565

2116 .CATTTGAGCAATCACGCCCAAGCAGTTTTTGGTGTCGCACCGCATCAAA 2164
      :::::::::|||:::                ||||||       ::::::|
 565 eAsnIleThrAsnGlnIleThrIleLysAlaGlyValAspGlyGluAsnS 582

2165 GCCACACAATCTGTACACGTTCGGACTGGACGGGTCTGACAAATTGTGTC 2214
     ||                        ||||||   |||:::   :::|||   :::
 582 erAsp...............SerAspAlaThrAsnAsnAlaAsnLeuThr 593

2215 GAAAAAACC............ATTACCGACGATAAAGTGATTGCTTCATT 2252
        ||||||             :::|||:::|||       |||::::::::
 594 IleLysThrLysGluLeuLysLeuThrGlnAspLeuAsnIleSerGlyPh 610

2253 GACTAAGACCGACATCAGC..........................GGCA 2275
     ::::|||:::::::|||:::                         ||||
 610 eAsnLysAlaGluIleThrAlaLysAspGlySerAspLeuThrIleGlyA 627

2276 ATGTCGATCTTGCCGATCACGCTCATTTAAATCTCACAGGGCTT...... 2319
     ||::::::   ||||||   ::::::   :::   :::   :::
 627 snThrAsnSerAlaAspSerThrAsnAlaLysLysValThrPheAsnGln 643

2320 ..................................GCCACACTCAACGG 2333
                                        :::|||||||||:::
 644 ValLysAspSerLysIleSerAlaGlyAspHisAsnValThrLeuAsnSe 660

2334 CAATCTTAGTGCAAATGGCGATACACGTTATACA................ 2367
     :::::::::::::::|||:::|||      |||
 660 rLysValGluThrSerGlyAsnThrAspAsnThrGlyAspGlySerGlyA 677

2368 ....................................GTCAGCCACAAC 2379
                                          |||::::::|||
 677 snAsnAlaGlyLeuThrIleAlaAlaLysAsnValGluValLysAsnAsn 693

2380 GCCACCCAAAACGGCAACCTTAGCCTCGTGGGCAATGCCCAAGCAACATT 2429
        |||:::|||   ::::::::::::::::::   :::   |||
 694 IleThrSerAsnLysThrValAsnIleThrAlaSerGluLysLeuThrTh 710

2430 TAAT...CAAGCCACATTAAACGGCAACACATCGGCTTCGGGCAATGCTT 2476
      :::   :::||||||:::|||::::::|||        ||||||:::
 710 rLysAlaAspAlaThrIleAsnAlaThrThr.........GlyAsnValG 724

2477 CATTTAATCTAAGCGACCACGCCGTACAAAACGGCAGTCTGACGCTTTCC 2526
     ::   :::   :::          :::::::::         :::   |||
 724 luValThrAlaLysThrGlyAspIleLysGlyGluValLysSerThrSer 740

2527 GGCAACGCTAAGGCAAACGTAAGCCATTCCGCACTCAAC...GGTAATGT 2573
     ||||||::::::   :::::::::   ::::::||||||   :::||||||
 741 GlyAsnValAsnIleThrAlaAsnGlyAspThrLeuAsnValSerAsnVa 757

2574 CTCC..................CTAGCCGATAAGGCAGTATTCCATTTTG 2605
     ||||                  |||||||||:::   :::       :
 757 lSerGlyAsnAlaValThrIleThrAlaAspLysGlyLysLeuThrThrG 774
```

```
2606 AAAGCAGCCGCTTTACCGGACAAATCAGCGGCGGCAAGGATACGGCATTA 2655
     :::::|||        ::::::        |||::::::   :::|||:::|||
 774 lnAlaSer......SerSerIleThrSerAsnAsnGlyGlnThrThrLeu 788

2656 CACTTAAAAGACAGCGAATGGACGCTGCCGTCAGGCACGGAATTAGGCAA 2705
            ||||||                       |||:::    |||::
 789 ThrAlaLysAsp....................GlySerIleAlaGlySe 798

2706 TTTAAACCTTGACAACGCCACCATTACACTCAATTCCGCCTATCGCCACG 2755
     ::::|||         |||::::::|||||||||:::
 798 rIleAsnAla......AlaAsnValThrLeuAsnThr............. 808

2756 ATGCGGCAGGGGCGCAAACCGGCAGTGCGACAGATGCGCCGCGCCGCCGT 2805
                     ||||||:::
 809 ................ThrGlyThr........................ 811

2806 TCGCGCCGTTCGCGCCGTTCCCTATTATCCGTTACACCGCCAACTTCGGT 2855
                             |||::::::::    ::::::::
 812 .......................LeuThrThrValGluGlySerAsnIl 820

2856 AGAATCCCGTTTCAACACGCTGACGGTAAACGGCAAATTGAACGGTCAGG 2905
     ::::::       :::||||||::::::|||:::|||
 820 eAsnAlaAlaSerGlyThrLeuValIleAsnAlaLys............. 832

2906 GAACATTCCGCTTTATGTCGGAACTCTTCGGCTACCGCAGCGACAAATTG 2955
                                         :::   ||||||
 833 ...................................AspAlaLysLeu 836

2956 AAGCTGGCGGAAAGTTCCGAAGGCACTTACACCTTGGCGGTCAACAATAC 3005
     :::   |||   |||::::::   |||   :::   |||:::|||   ::
 837 AsnGlyAlaAlaSerGlyAspHisThrValValAsnAlaThrAsnAlaSe 853

3006 CGGCAACGAACCTGCAAGCCTCGAACAATTGACGGTAGTGGAAGGAAAAG 3055
     :|||:::                                  |||:::
 853 rGlySer..................................GlySerV 858

3056 ACAACAAACCGCTGTCCGAAAACCTTAATTTCACCCTGCAAAACGAACAC 3105
       :::         |||:::|||:::|||:::|||   :::   :::
 858 alThrAlaValThrSerSerAsnValAsnIleThrGlyAspLeuSerThr 874

3106 GTCGATGCCGGCGCGTGGCGTTACCAACTCATCCGCAAAGACGGCGAGTT 3155
     |||::::::               ::::::|||   |||:::|||
 875 ValAsnGly...........LeuAsnIleIleSerLysAsnGly..... 885

3156 CCGCCTGCATAATCCGGTCAAAGAACAAGAGCTTTCCGACAAACTCGGCA 3205
             :::|||   |||                     |||
 886 .......ArgAsnThrVal....................ValLeuLysG 893

3206 AGGCAGAAGCCAAAAAACAGGCGGAAAAAGACAACGCGCAAAGCCTTGAC 3255
       :::|||   :::   :::      :::   ::::::   |||::::::
 893 lyThrGluIleGluValLysTyrIleGlnProGlyValAlaSerValGlu 909

3256 GCGCTGATTGCGGCCGGGCGCGATGCCGTCGAAAAGACAGAAAGCGTTGC 3305
        :::|||   |||   |||   ::::::|||||||:::::::::::::::
 910 GluValIleGluAlaLysArg...ValLeuGluLysValLysAspLeuSe 925

3306 CGAACCGGCCCGGCAGGCAGGCGGGGAAAATGTCGGCATTATGCAGGCGG 3355
     ::::      |||::::::   ::::::   :::|||:::      :::
 925 rAspGluGluArgGluThrLeuAlaLys...LeuGlyValSerAlaVal. 940

3356 AGGAAGAGAAAAAACGGGTGCAGGCGGATAAAGACACCGCCTTGGCGAAA 3405
                 :::   ::::::      ::::::
 941 ..........ArgPheIleGluProAsnAsnThrIleThrValAsnThr 953

3406 CAGCGCGAAGCGGAAACCCGGCCGGCTACC 3435
     |||:::|||      |||||||||::::::
 954 GlnAsnGluPheThrThrArgProSerSer 963
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01842

seq_documentation_block:

```
ID   AAB01842 standard; Protein; 998 AA.
XX
AC   AAB01842;
XX
DT   11-SEP-2000  (first entry)
XX
DE   Haemophilus influenzae strain 15 HMW1A protein, SEQ ID NO:59.
XX
KW   HMW protein; hmw gene; hmwA1; hmwA2; high molecular weight;
KW   non-typeable Haemophilus influenzae; NTHi; non-encapsulated;
KW   recombinant production; Escherichia coli; antibacterial; vaccine;
KW   human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;
KW   detection; diagnosis.
XX
OS   Haemophilus influenzae strain 15.
XX
PN   WO200020609-A2.
XX
PD   13-APR-2000.
XX
PF   07-OCT-1999;   99WO-CA00938.
XX
PR   07-OCT-1998;   98US-0167568.
PR   08-DEC-1998;   98US-0206942.
XX
PA   (CONN-) CONNAUGHT LAB LTD.
XX
PI   Loosmore SM,  Yang Y,  Klein MH;
XX
DR   WPI; 2000-303789/26.
DR   N-PSDB; AAA52191.
XX
PT   Nucleic acid molecule for producing recombinant high molecular weight
PT   proteins of Haemophilus which are used as a vaccine to provide
PT   protection against Haemophilus induced diseases in humans -
XX
PS   Claim 12; Fig 26A-O; 307pp; English.
XX
CC   The invention relates to the recombinant production of Haemophilus
CC   influenzae high molecular weight (HMW) proteins in Escherichia coli. The
CC   expression construct used to effect recombinant expression comprises a
CC   promoter functional in E. coli (e.g., the T7 promoter) operably linked
CC   to a modified hmwABC operon from a non-typeable (non-encapsulated) H.
CC   influenzae (NTHi). Most HMW-expressing NTHi strains contain two hmw gene
CC   clusters termed hmw1ABC and hmw2ABC. Each hmwABC operon comprises hmwA,
CC   hmwB and hmwC genes. The hmwA genes encode the structural HMWA proteins
CC   and the hmwB and hmwC genes encode accessory proteins which are
CC   responsible for post-translational processing and secretion of the HMWA
CC   proteins. The modified hmwABC operon used in the expression construct of
CC   the invention contains an A gene modified such that it encodes only the
CC   mature HMWA. The invention also discloses hmwA genes (AAA52175-A52198)
CC   and HMWA proteins (AAB01824-B01849) from the non-typeable H. influenzae
CC   strains Joyc, K1, K21, LCDC2, PMH1, 15 and 12. The nucleic acids and
CC   vectors are used for the production of recombinant H. influenzae HMW
CC   proteins which can be used as vaccines to mediate a humoral or
CC   cell-mediated immune response to provide protection against diseases in
CC   humans caused by H. influenzae (e.g., otitis media, epiglottitis,
CC   pneumonia and tracheobronchitis). The HMW proteins are also useful as
CC   antigens in immunoassays for detecting antibodies against Haemophilus,
CC   HMW proteins and/or HMW peptides. The nucleotide sequences encoding the
CC   HMW proteins can be used to isolate and clone hmw genes from other
CC   non-typeable strains of Haemophilus via hybridisation reactions. The
CC   present sequence represents an HMWA protein from a non-typeable strain of
CC   H. influenzae.
XX
SQ   Sequence   998 AA;
```

alignment_scores:

```
           Quality:  261.00          Length:   1160
             Ratio:  0.476             Gaps:     58
 Percent Similarity:  47.241   Percent Identity: 19.914
```

alignment_block:

```
US-09-303-518D-649 x AAB01842   ..

Align seg 1/1  to: AAB01842  from: 1  to: 998

 508 GATGCAGAACCTGTTGAAATGACCAGTTATATGGATGGGCGGAAATATAT 557
     |||      :::|||    ||||||     |||:::       ||
   8 AspAsnValThrIleGluAlaProSerTyrSerArgGlyAsnAlaGlyIl 24

 558 CGATCAAAATAATTACCCTGACCGTGTTCGTATTGGGGCAGGC...AGGC 604
     ||||   :::::::|||                |||:::|||   ::::
  24 eAsp...SerGluPheProGly...........GlySerGlyThrLysG 36

 605 AATATTGGCGATCTGATGAAGATGAGCCC.........AATAACCGCGAA 645
     ::      :::::::::   ::::::|||         :::|||:::
  36 luSerProLysThrAsnGlyGluGlnProThrValLeuThrAsnGluThr 52

 646 AGTTCATATCATATTGCAAGTGCGTATTCTTGGCTCGTT........... 684
       |||    ::::::   |||:::   :::||||:::::
  53 IleSerAsnTyrLeuLysSerGly...ThrTrpValMetAsnIleThrAl 68

 684 .................................................. 684

  68 aLysLysAsnLeuThrValAsnSerSerIleAsnIleGlyAspSerSerH 85

 684 .................................................. 684

  85 isLeuIleLeuHisSerGluGlyLysAsnAsnGlyGlyValLysIleLys 101

 685 ..................GGTGGCAATACCTTTGCACAAAATGGATCA.. 714
                       |||||||||      |||:::|||:::
 102 GluAspIleThrSerAsnGlyGlyAsnLeuThrIleGlnSerGlyGlyTr 118

 715 ...........................GGTGGTGGCACAGTCAACTTA. 735
                                |||   ||||||:::|||:::
 118 pValAspValHisLysAsnIleThrLeuGlyThrGlyThrLeuAsnIleT 135

 736 ...............................GGTAGTGAAAAAATTAAA 753
                                   |||:::||||||   :::
 135 hrAlaLysGlySerIleAlaPheGluGlyAsnGlyThrGluLysAlaArg 151

 754 CATAGCCCATATGGTTTTTTACCAACAGGAGGCTCATTTGGCGACAGTGG 803
     ::::::     :::   :::   :::   |||:::::::   ::::::||
 152 AsnAlaSerSerAlaGlnIleThrAlaGlnGlyThrIleThrAsnThrGl 168

 804 CTCACCAATGTTTATCTATGATGCCCAAAAGCAAAAGTGGTTA....... 846
     |:::                       |||||||||      |||
 168 yAsp.....................GlnLysGlnLeuArgLeuAsnAsnV 178

 847 .....ATTAATGGGGTATTGCAAACGGGCAACCCCTATATAGGAAAAAGC 891
          |||||||||        ||||||         ||||||   :::
 178 alSerIleAsnGly.........ThrGly.........IleGlyLeuAsn 188

 892 AATGGCTTCCAGCTGGTTCGTAAAGATTGGTTCTATGATGAAATCTTTGC 941

 188 .................................................. 188

 942 TGGAGATACCCATTCAGTATTCTACGAACCACGTCAAAATGGGAAATACT 991
                       |||   :::   :::   |||   ::::::
 189 ..................PheValSerIleGlnProAsnThrSerHisA 199

 992 CTTTTAACGACGATAATAATGGCACAGGAAAAATCAATGCCAAACATGAA 1041
        |||:::   :::        :::|||:::::         |||
 199 rgPheAspGlyGluLeuIleIleSerGlyArgVal.........HisVal 212

1042 CACAATTCTCTGCCTAATAGATTA.........AAAACACGAACCGTTCA 1082
     :::::::::   |||:::::|||         |||:::         ::
 213 AsnGlnThrThrProLysAsnLeuSerPheTrpLysValSerAspGluSe 229

1083 ATTGTTTAATGTTTCTTTATCCGAGACAGCAAGAGAACCTGTTTATCATG 1132
     :   :::|||||||||         |||:::::::|||           :
 229 rTyrTrpAsnValSer...HisLeuThrValLysGlu.........LysS 242

1133 CTGCAGGTGGTGTCAACAGTTATCGACCCAGACTGAATAATGGAGAAAAT 1182
         ::|||   :::   :::::::::       :::   |||:::|||::::::
 242 erAlaPheSerPheThrLysPheAlaLeuAsnAsnAsnHisGlyArgGlu 258

1183 ATTTCCTTTATTGACGAAGGAAAAGGCGAATTGATACTTACCAGCAACAT 1232
                                          |||||||:::
 259 ......................................ThrSerArgTy 262

1233 CAATCAAGGTGCTGGAGGATTATATTTCCAAGGA...............G 1267
       ::::::|||   ||||||:::   |||::::::              :
 262 rArgLysGly...GlyGlyValIlePheArgSerProThrGlyHisThrA 278

1268 ATTTTACGGTCTCGCCTGAAAATAACGAAACTTGGCAAGGCGCGGGCGTT 1317
     ::||||||||:::                      ||||||:::   :::
 278 snPheThrValLys....................GlnGlySerValAla 287

1318 CATATCAGT............GAAGACAGTACCGTTACTTGGAAAGTAAA 1355
     :::::|||            :::||||:::::   :::   ::::::
 288 AsnPheSerPheLysAlaLysAsnAspThrAsnHisAlaAsnGlnLeuPr 304

1356 CGGCGTGGCAAACGACCGCCTGTCCAAAATCGGCAAAGGCACGCTG.... 1401
                 |||::::::::|||      |||   |||   :::
 304 oIleGlnPheAsnSerAsnIleSerValAspGlyGlyGlyLysValLeuP 321

1402 ...................CACGTTCAAGCCAAAGGGGAAAACCAAGGC 1431
                             ::::::   |||   :::::::::
 321 heCysIleThrSerAsnTyrSerGlyArgSerValGlyIleGlyMetSer 337

1432 TCGATCAGCGTGGGCGACGGTACAGTCATTTTGGATCAGCAGGCAGACGA 1481
     ||||||:::|||:::||||||:::   :::        :::::::::
 338 SerIleAsnValSerAspGlySerAsnLeuThrPheAsnSerSer..... 352

1482 TAAAGGCAAAAAACAAGCCTTTAGTGAAATCGGCTTGGTCAGCGGCAGGG 1531
                                          :::   |||:::
 353 ...................................IleArgGlyGlnG 357

1532 GTACGGTGCAACTGAATGCCGATAATCAGTTCAACCCCGACAAACTCTAT 1581
        :::   :::::::::   |||      :::|||              :::
 357 luAlaPheAsnIleSerLysAspLeuThrIleAsnAlaThrGlySerPhe 373

1582 TTCGGCTTTCGCGGCGGACGTTTGGATTTAAACGGGCATTCGCTTTCGTT 1631
     |||   :::               :::||||||::::::::::::
 374 PheGluLeuGlyGlnTyrSerAspThrPheAsnGlyAsnGlyPheAsnHi 390

1632 CCACCGTATTCAAAATACCGATGAA............GGGGCGATGATTG 1669
              |||:::::::|||   :::            |||:::   ::::
 390 sAspAlaIleLysSerThrHisAsnIleSerIleLeuGlyGlyAsnValT 407

1670 TCAACCACAATCAAGACAAAGAATCCACCGTTACCATTACAGGCAATAAA 1719
     ::      :::||||||::::::|||      ||||||||||||||
 407 hrLeuGlyGlyGlnAspSerSerSer......ThrIleThrGlyAsnIle 421

1720 GATATTGCTACAACCGGCAATAACAACAGCTTGGATAGCAAAAAAGAAAT 1769
     :::|||:::   ::::::|||   :::
 422 AsnIleSerGlnAlaAlaAsnValThrLeuArg................. 432

1770 TGCCTACAACGGTTGGTTTGGCGAGAAAGATACGACCAAAACGAACGGGC 1819
      ||||||||||||         :::|||:::   |||      |||
 433 .AlaTyrAsnGlyAsnGlyArgAsnLysGlnLeuThrLeuGlyAsn.... 447

1820 GGCTCAACCTTGTTTACCAGCCCGCCGCAGAAGACCGCACCCTGCTGCTT 1869
                         :::   |||   :::      |||:::
 448 ...................ValSerIleGluGlyAsnLeuSerLeuIle 457

1870 TCCGGCGGAACAAATTTAAACGGCAACATCACG...CAAACAAACGGCAA 1916
     :::::::::::|||:::|||||||||::::::   :::   |||:::||
 458 GlyAlaSerAlaAsnIleAsnGlyAsnLeuSerValLysGluAsnAlaLy 474

1917 ACTGTTTTTCAGCGGCAGACCA.................ACACCGCACG 1948
     |      |||:::|||:::                   |||
 474 s......PheLysGlyGluThrGlnAspAsnLeuAsnIleThrGlyThrP 489

1949 CCTACAATCATTTAAACGACCATTGGTCGCAAAAAGAGGGCATTCCTCGC 1998
        |||:::   ::::::      :::   ::::::|||:::   :::
```

```
 489 heIleAsnAsnGlyAspSerLysIleAsnIleSerGlnGlyValValLys 505

1999 ...GGGGAAATCGTGTGGGACAACGACTGGATCAACCGCACATTTAAAGC 2045
        |||::::::::   |||:::|||   :::|||   |||       ||
 506 LeuGlyAsnValThrAsnAspGlyAsp...LeuAsnIleThrThrHisAl 521

2046 GGAAAACTTCCAA......ATTAAAGGCGGACAGGCGGTG.......... 2079
     |::::::   |||        |||   ||||||:::    :::
 521 aLysHisAsnGlnArgSerIleIleGlyGlyAspIleIleAsnLysLysG 538

2080 .........................................GTTTCC 2085
                                              ::::::
 538 lySerLeuAsnIleThrAspSerAsnLysAsnAlaGluIleGlnIleGly 554

2086 CGCAATGTTGCCAAAGTGAAAGGCGATTGG................... 2115
        |||::::::::    :::|||:::
 555 GlyAsnIleSerGlnLysGluGlyAsnLeuThrIleSerSerAspLysIl 571

2116 .CATTTGAGCAATCACGCCCAAGCAGTTTTTGGTGTCGCACCGCATCAAA 2164
      ::::::::|||:::                 ||||||       ::::::|
 571 eAsnIleThrAsnGlnIleThrIleLysAlaGlyValAspGlyGluAsnS 588

2165 GCCACACAATCTGTACACGTTCGGACTGGACGGGTCTGACAAATTGTGTC 2214
     ||                  ||||||   |||:::   :::|||    :::
 588 erAsp..............SerAspAlaThrAsnAsnAlaAsnLeuThr 599

2215 GAAAAAACC...........ATTACCGACGATAAAGTGATTGCTTCATT 2252
       ||||||                :::|||:::|||      |||::::::::
 600 IleLysThrLysGluLeuLysLeuThrGlnAspLeuAsnIleSerGlyPh 616

2253 GACTAAGACCGACATCAGC.........................GGCA 2275
     ::::|||::::::|||:::                           ||||
 616 eAsnLysAlaGluIleThrAlaLysAspGlySerAspLeuThrIleGlyA 633

2276 ATGTCGATCTTGCCGATCACGCTCATTTAAATCTCACAGGGCTT...... 2319
     ||::::::   ||||||   ::::::   :::   :::   :::
 633 snThrAsnSerAlaAspSerThrAsnAlaLysLysValThrPheAsnGln 649

2320 ..................................GCCACACTCAACGG 2333
                                       :::|||||||||::
 650 ValLysAspSerLysIleSerAlaGlyAspHisAsnValThrLeuAsnSe 666

2334 CAATCTTAGTGCAAATGGCGATACACGTTATACA................ 2367
     :::::::::::::::|||:::|||      |||
 666 rLysValGluThrSerGlyAsnThrAspAsnThrGlyAspGlySerGlyA 683

2368 ....................................GTCAGCCACAAC 2379
                                          |||::::::|||
 683 snAsnAlaGlyLeuThrIleAlaAlaLysAsnValGluValLysAsnAsn 699

2380 GCCACCCAAAACGGCAACCTTAGCCTCGTGGGCAATGCCCAAGCAACATT 2429
        |||:::|||    ::::::::::::::::::   :::   |||
 700 IleThrSerAsnLysThrValAsnIleThrAlaSerGluLysLeuThrTh 716

2430 TAAT...CAAGCCACATTAAACGGCAACACATCGGCTTCGGGCAATGCTT 2476
     :::    :::||||||:::|||:::::::|||          |||||||::::
 716 rLysAlaAspAlaThrIleAsnAlaThrThr.........GlyAsnValG 730

2477 CATTTAATCTAAGCGACCACGCCGTACAAAACGGCAGTCTGACGCTTTCC 2526
     ::    :::   :::           :::::::::       :::    |||
 730 luValThrAlaLysThrGlyAspIleLysGlyGluValLysSerThrSer 746

2527 GGCAACGCTAAGGCAAACGTAAGCCATTCCGCACTCAAC...GGTAATGT 2573
     ||||||::::::   :::::::::   ::::::||||||    :::|||||
 747 GlyAsnValAsnIleThrAlaAsnGlyAspThrLeuAsnValSerAsnVa 763

2574 CTCC.................CTAGCCGATAAGGCAGTATTCCATTTTG 2605
     ||||                    |||||||||:::   :::      :
 763 lSerGlyAsnAlaValThrIleThrAlaAspLysGlyLysLeuThrThrG 780

2606 AAAGCAGCCGCTTTACCGGACAAATCAGCGGCGGCAAGGATACGGCATTA 2655
     ::::::|||        ::::::      |||::::::   :::|||:::|||
 780 lnAlaSer......SerSerIleThrSerAsnAsnGlyGlnThrThrLeu 794

2656 CACTTAAAAGACAGCGAATGGACGCTGCCGTCAGGCACGGAATTAGGCAA 2705
        ||||||                         |||:::       |||::
 795 ThrAlaLysAsp...................GlySerIleAlaGlySe 804

2706 TTTAAACCTTGACAACGCCACCATTACACTCAATTCCGCCTATCGCCACG 2755
     :::::|||          |||:::::::|||||||||:::
 804 rIleAsnAla......AlaAsnValThrLeuAsnThr............ 814

2756 ATGCGGCAGGGGCGCAAACCGGCAGTGCGACAGATGCGCCGCGCCGCCGT 2805
                       |||||||:::
 815 ................ThrGlyThr...................... 817

2806 TCGCGCCGTTCGCGCCGTTCCCTATTATCCGTTACACCGCCAACTTCGGT 2855
                            |||::::::::::       ::::::::
 818 ......................LeuThrThrValGluGlySerAsnIl 826

2856 AGAATCCCGTTTCAACACGCTGACGGTAAACGGCAAATTGAACGGTCAGG 2905
     ::::::       :::||||||::::::|||:::|||
 826 eAsnAlaAlaSerGlyThrLeuValIleAsnAlaLys............ 838

2906 GAACATTCCGCTTTATGTCGGAACTCTTCGGCTACCGCAGCGACAAATTG 2955
                                         :::    ||||||
 839 ..................................AspAlaLysLeu 842

2956 AAGCTGGCGGAAAGTTCCGAAGGCACTTACACCTTGGCGGTCAACAATAC 3005
     :::   |||   |||:::::::   |||     :::    |||:::|||    ::
 843 AsnGlyAlaAlaSerGlyAspHisThrValValAsnAlaThrAsnAlaSe 859

3006 CGGCAACGAACCTGCAAGCCTCGAACAATTGACGGTAGTGGAAGGAAAAG 3055
     :|||:::                                    |||:::
 859 rGlySer................................GlySerV 864

3056 ACAACAAACCGCTGTCCGAAAACCTTAATTTCACCCTGCAAAACGAACAC 3105
        :::              |||:::|||:::|||:::|||    :::    :::
 864 alThrAlaValThrSerSerAsnValAsnIleThrGlyAspLeuSerThr 880

3106 GTCGATGCCGGCGCGTGGCGTTACCAACTCATCCGCAAAGACGGCGAGTT 3155
     |||::::::              :::::::|||    |||:::|||
 881 ValAsnGly...........LeuAsnIleIleSerLysAsnGly..... 891

3156 CCGCCTGCATAATCCGGTCAAAGAACAAGAGCTTTCCGACAAACTCGGCA 3205
             :::|||    |||                          |||
 892 .......ArgAsnThrVal...................ValLeuLysG 899

3206 AGGCAGAAGCCAAAAAACAGGCGGAAAAAGACAACGCGCAAAGCCTTGAC 3255
        :::|||    :::    :::     ::::::    |||::::::
 899 lyThrGluIleGluValLysTyrIleGlnProGlyValAlaSerValGlu 915

3256 GCGCTGATTGCGGCCGGGCGCGATGCCGTCGAAAAGACAGAAAGCGTTGC 3305
         :::|||   |||   |||    ::::::||||||:::::::::::::
 916 GluValIleGluAlaLysArg...ValLeuGluLysValLysAspLeuSe 931

3306 CGAACCGGCCCGGCAGGCAGGCGGGGAAAATGTCGGCATTATGCAGGCGG 3355
     ::::         |||::::::    ::::::    :::|||:::       :::
 931 rAspGluGluArgGluThrLeuAlaLys...LeuGlyValSerAlaVal. 946

3356 AGGAAGAGAAAAAACGGGTGCAGGCGGATAAAGACACCGCCTTGGCGAAA 3405
                 :::   ::::::   ::::::       ::::::
 947 ..........ArgPheIleGluProAsnAsnThrIleThrValAsnThr 959

3406 CAGCGCGAAGCGGAAACCCGGCCGGCTACC 3435
     |||:::|||      |||||||||::::::
 960 GlnAsnGluPheThrThrArgProSerSer 969
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01833

seq_documentation_block:
ID   AAB01833 standard; Protein; 1005 AA.
XX
AC   AAB01833;
XX
DT   11-SEP-2000 (first entry)

```
XX
DE   Haemophilus influenzae strain K21 mature HMW2A protein, SEQ ID NO:41.
XX
KW   Mature HMW protein; hmw gene; hmwA1; hmwA2; high molecular weight;
KW   non-typeable Haemophilus influenzae; NTHi; non-encapsulated;
KW   recombinant production; Escherichia coli; antibacterial; vaccine;
KW   human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;
KW   detection; diagnosis.
XX
OS   Haemophilus influenzae strain K21.
XX
PN   WO200020609-A2.
XX
PD   13-APR-2000.
XX
PF   07-OCT-1999;   99WO-CA00938.
XX
PR   07-OCT-1998;   98US-0167568.
PR   08-DEC-1998;   98US-0206942.
XX
PA   (CONN-) CONNAUGHT LAB LTD.
XX
PI   Loosmore SM,  Yang Y,  Klein MH;
XX
DR   WPI; 2000-303789/26.
DR   N-PSDB; AAA52182.
XX
PT   Nucleic acid molecule for producing recombinant high molecular weight
PT   proteins of Haemophilus which are used as a vaccine to provide
PT   protection against Haemophilus induced diseases in humans -
XX
PS   Claim 8; Fig 21A-O; 307pp; English.
XX
CC   The invention relates to the recombinant production of Haemophilus
CC   influenzae high molecular weight (HMW) proteins in Escherichia coli. The
CC   expression construct used to effect recombinant expression comprises a
CC   promoter functional in E. coli (e.g., the T7 promoter) operably linked
CC   to a modified hmwABC operon from a non-typeable (non-encapsulated) H.
CC   influenzae (NTHi). Most HMW-expressing NTHi strains contain two hmw gene
CC   clusters termed hmw1ABC and hmw2ABC. Each hmwABC operon comprises hmwA,
CC   hmwB and hmwC genes. The hmwA genes encode the structural HMWA proteins
CC   and the hmwB and hmwC genes encode accessory proteins which are
CC   responsible for post-translational processing and secretion of the HMWA
CC   proteins. The modified hmwABC operon used in the expression construct of
CC   the invention contains an A gene modified such that it encodes only the
CC   mature HMWA. The invention also discloses hmwA genes (AAA52175-A52198)
CC   and HMWA proteins (AAB01824-B01849) from the non-typeable H. influenzae
CC   strains Joyc, K1, K21, LCDC2, PMH1, 15 and 12. The nucleic acids and
CC   vectors are used for the production of recombinant H. influenzae HMW
CC   proteins which can be used as vaccines to mediate a humoral or
CC   cell-mediated immune response to provide protection against diseases in
CC   humans caused by H. influenzae (e.g., otitis media, epiglottitis,
CC   pneumonia and tracheobronchitis). The HMW proteins are also useful as
CC   antigens in immunoassays for detecting antibodies against Haemophilus,
CC   HMW proteins and/or HMW peptides. The nucleotide sequences encoding the
CC   HMW proteins can be used to isolate and clone hmw genes from other
CC   non-typeable strains of Haemophilus via hybridisation reactions. The
CC   present sequence represents a mature HMWA protein from a non-typeable
CC   strain of H. influenzae.
XX
SQ   Sequence   1005 AA;


alignment_scores:
           Quality:  259.50        Length:   1159
             Ratio:  0.451           Gaps:     54
 Percent Similarity:  49.698   Percent Identity: 20.276

alignment_block:
 US-09-303-518D-649 x AAB01833   ..

 Align seg 1/1  to: AAB01833  from: 1  to: 1005
```

```
 226 AAAAAAGGGGAGTTGGTCGGCAAATCAATGACAAAAGCCCCGATGATTGA 275
     :::||||||      |||:::    |||   ||||||         :::::
  45 GlnLysGlyIleGluValAsnIleSerAlaThrLysAsnValThrValAs 61

 276 TTTTTCTGTGGTGTCGCGTAACGGCGTGGCGGCATTGGTGGGCGATCAAT 325
     :   :::|||      :::|||||:::   :::|||    :::::::::
  61 nAlaAspValAspValLysAsnGlyThrLeuValLeuHisSerGlnArgA 78

 326 ATATTGTGAGCGTGGCACATAACGGCGGCTATAACAACGTTGATTTTGGT 375
           |||:::::        ||||||:::  ||||||:::
  78 snGlyValLysIle......AsnGlyAsn.IleThrSerThrGln..... 90

 376 GCGGAAGGAAGAAATCCCGATCAACATCGTTTTACTTATAAAATTGTGAA 425
                                                     |
  91 .................................................A 91

 426 ACGGA...ATAATTATAAAGCAGGGACTAAAGGCCATCCTTATGGCGGCG 472
     |||||    :::    ||||||   |||   |||:::    :::
  91 snGlyAsnLeuThrIleLysThrGlyGlyLysValAspValHisLysAsn 107

 473 ATTATCATATGCCGCGTTTGCATAAATTTGTCACAGATGCAGAACCTGTT 522
     |||   :::            :::|||::::::::    :::|||:::
 108 IleThrLeuGlyMetGlyPheLeuAsnIleThrSerAspAsnAsnIle.. 123

 523 GAAATGACCAGTTATATGGATGGGCGGAAATATATCGATCAAAATAATTA 572
                :::::      |||    |||::::::|||
 124 .........ThrPheGluLysGlyAspAsnLeuThrIleThr....... 134

 573 CCCTGACCGTGTTCGTATTGGGGCAGGCAGGCAATATTGGCGATCTGATG 622
                                ||||||||      |||::::
 135 ......................AlaGlnGlyAsnIle..IleSerAsnG 143

 623 AAGATGAGCCCAATAACCGCGAAAGTTCATATCATATTGCAAGTGCGTAT 672
     ::::::::    :::    |||    |||:::       :::
 143 lnGluAsnLysGlnLeuArgPheSerAsnValSerLeu............ 155

 673 TCTTGGCTCGTTGGTGGCAATACCTTTGCACAAAATGGATCAGGT..... 717
     :::    :::    :::|||    ||||||:::    |||    ::::::
 156 AsnGlyMetGlyAlaGlyLeuThrPheThrAlaAsnLysGlyAsnHisTh 172

 718 .........GGTGGCACAGTCAACTTAGGTAGTGAAAAAATT......A 751
                  ||||||:::|||::::::::::::    :::       :
 172 rHisLysPheAspGlyThrLeuAsnIleSerGlyLysValValIleAsnG 189

 752 AACATAGCCCATATGGTTTTTTACCAACAGGAGGCTCA............ 789
     ::    :::|||::::::::    |||    :::::::|||
 189 lnThrThrProHisAsnIleAlaProTrpAsnAlaSerAlaAspSerTyr 205

 790 ....................TTTGGCGACAGTGGCTCACCAATGTTTAT 818
                         :::|||::::::::::::       |||||
 206 TrpAsnValThrThrLeuThrLeuGlyAsnAsnAlaGlnPheThrPheIl 222

 819 C......TATGATGCCCAAAAGCAAAAGTGGTTAATTAATGGGGTATTGC 862
     |          |||::::::::::::::        |||    :::|||    :
 222 eLysPheValAspSerAsnArgSerValAlaLeuAsnSerGlySerArgS 239

 863 AA...ACGGGCAACCCCTATATAGGAAAAAGCAATGGCTTCCAGCTGGTT 909
     ::    :::|||      :::    ||||||:::|||    :::
 239 erPheAlaGlyValLysPheTyrGlyLysAsnAsnGluMet......... 252

 910 CGTAAAGATTGGTTCTATGATGAAATCTTTGCTGGAGATACCCATTCAGT 959
                         ::::::      ||||||:::    :::||
 253 ....................LysPheAsnIleGlyAspAsnAlaAsnVa 262

 960 ATTCTACGAACCACGTCAAAATGGGAAATACTCTTTTAACGAC....... 1002
     |    ::::::    ::::::|||    :::    |||    |||
 262 lGluPheLysLeuLysSerAsnAspAsnThrSerAsnAsnLysProLeuP 279

1003 ............................GATAATAATGGCACAGGAAAA 1023
                                 :::|||||||||    :::
 279 roIleGlnPheLeuSerAsnIleSerAlaThrGlyAsnGlyThrValSer 295

1024 ATCAATGCCAAACATGAACACAATTCTCTGCCTAATAGATTAAAAACACG 1073
```

```
        ::::::               |||   :::|||      |||
 296 PheAspIle.........HisAlaAsnLeuSerAlaArgSerThrGluLe 309

1074 AACCGTTCAATTGTTTAATGTTTCT......................... 1098
       ::::::::|||:::|||:::|||
 309 uAsnMetSerLeuIleAsnIleSerAsnGlyValAsnPheSerIleAsnS 326

1099 ...................TTATCCGAGACAGCAAGAGAACCTGTTTAT 1128
                          |||      ::::::   :::
 326 erHisValArgGlyAsnAsnAlaPheGluIleLysLysAspLeuIleIle 342

1129 CATGCTGCAGGTGGTGTCAACAGTTATCGACCCAGACTGAATAATGGAGA 1178
      :::|||:::|||:::      :::   :::          ::::::   ::
 343 AsnAlaThrGlySerAsnPheAsnLeuLysGlnThrLysAspLysPheAs 359

1179 AAATATTTCCTTTATTGACGAAGGAAAAGGCGAATTGATACTTACCAGCA 1228
      :|||             :::     |||:::   ::::::   |||    |
 359 pAsn............SerTyrGluLysAsnAlaIlePheSerThrHisA 372

1229 ACATCAATCAAGGTGCTGGAGGATTATATTTCCAAGGAGATTTTACGGTC 1278
      ||::::::                  :::   |||:::   |||:::
 372 snLeuThr..................IleLeuGlyGlyAsnValThrLeu 382

1279 TCGCCTGAAAATAACGAAACTTGGCAAGGCGCGGGCGTTCATATCAGTGA 1328
      :::   ||||||:::::::::        :::::::::::|||::::::
 383 GlyGlyGluAsnSerSerSerAsnIleLysGlyAsnIleAsnIleAsnSe 399

1329 AGACAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAAACGACCGCCTGT 1378
      :    ::::::||||||    ::::::::::::
 399 rLysAlaAsnValThrLeuGlnAlaHisAla................... 409

1379 CCAAAATCGGCAAAGGCACGCTGCACGTTCAAGCCAAAGGGGAA...... 1422
                    ||||||   |||::::::   |||    :::
 410 ..............GlyThrSerHisLeuAspLysLysGluArgThrLeu 421

1423 AACCAAGGCTCGATCAGCGTGGGCGACGGTACAGTCATTTTGGATCAGCA 1472
      :::    |||:::::::||||||||||   :::        |||:::   :::::
 422 ThrLeuGlyAsnValSerValGlyGlyAsnLeuAsnIleIleGlySerAs 438

1473 GGCAGACGATAAAGGCAAAAAACAAGCCTTTAGTGAAATCGGCTTGGTCA 1522
      :|||         |||   :::   :::::::::|||   :::          :
 438 nAlaHisIleAspGly...AsnLeuSerIleAlaGluSerAlaLysPheG 454

1523 GCGGCAGGGGTACGGTGCAACTGAATGCCGATAATCAGTTCAACCCCGAC 1572
      ::|||:::   :::   :::||||||        :::   |||:::
 454 lnGlyLysThrAsnAsnAsnLeuAsnIleThrGlyThrPheThr...... 468

1573 AAACTCTATTTCGGCTTTCGCGGCGGACGTTTGGATTTAAACGGGCATTC 1622
                          ::::::|||      |||:::|||
 469 .................AsnAsnGlyThrAlaAspIleAsn........ 476

1623 GCTTTCGTTCCACCGTATTCAAAATACCGATGAAGGGGCGATGATTGTCA 1672

 476 .................................................. 476

1673 ACCACAATCAAGACAAAGAATCCACCGTTACCATTACAGGCAATAAAGAT 1722
                   |||:::::::::|||    :::    |||
 477 ..........IleLysGlnGlyValValLysLeuGlnGly......... 486

1723 ATTGCTACAACCGGCAATAACAACAGCTTGGATAGCAAAAAAGAAATTGC 1772

 486 .................................................. 486

1773 CTACAACGGTTGGTTTGGCGAGAAAGATACGACCAAAACGAACGGGCGGC 1822
                              |||     |||     :::||||||:::|
 487 .........................AspIleThr...AsnAsnGlyAsnL 494

1823 TCAACCTTGTTTACCAGCCCGCCGCAGAAGACCGCACCCTGCTGCTTTCC 1872
      ||||::::::    :::    ::::::::::::::|||:::
 494 euAsnIleThrThrAsnAlaSerValAsnGlnLysThrIle.........507

1873 GGCGGAACAAATTTAAACGGCAACATCACGCAAACAAACGGCAAACTGTT 1922
                   :::||||||||||||||||:::    :::|||    |||

 508 ............IleAsnGlyAsnIleThrAsnLysLysGlyAspLeuAs 520

1923 TTTCAGCGGCAGACCAACACCGCACGCCTACAATCATTTAAACGACCATT 1972
      ::::::          :::    :::|||     ::::::::    :::
 520 nIleLysAspIleLysAla...AsnAlaGluIleGlnIleGlyGlyAsnI 536

1973 GGTCGCAAAAAGAGGGCATTCCTCGCGGGGAAATCGTGTGGGACAACGAC 2022
       |||||||||||               |||:::::::::   ::::::|||
 536 leSerGlnLysGlu...........GlyAsnLeuThrIleSerSerAsp 548

2023 TGGATCAACCGCACATTTAAAGCGGAAAACTTCCAAATTAAAGGCGGACA 2072
       ||||||    |||    :::    |||
 549 LysIleAsnIleThrLysArgIleGlu...................... 557

2073 GGCGGTGGTTTCCCGCAATGTTGCCAAAGTGAAAGGCGATTGGCATTTGA 2122
                                  :::|||:::|||          :
 558 ..........................IleLysAlaAsp.........T 562

2123 GCAATCACGCCCAAGCAGTTTTTGGTGTCGCACCGCATCAAAGCCACACA 2172
      ::::::::::::::::          |||||||||    :::    :::    |||
 562 hrAspGlnGlyAsnSerAspSerGlyValAlaSerAsnAlaAsnLeuThr 578

2173 ATCTGTACACGTTCGGACTGGACGGGTCTGACAAATTGTGTCGAAAAAAC 2222
      |||    |||:::                              |||    ||
 579 IleLysThrLys.............................GluLeuTh 585

2223 CATTACCGACGATAAAGTGATTGCTTCATTGACTAAGACCGACATCAGCG 2272
      |:::|||||||:::       |||::::::::::::|||:::::::|||:::
 585 rLeuThrAspAsnLeuAsnIleSerGlyPheAsnLysAlaGluIleThrA 602

2273 GC......AATGTCGATCTTGCCGATCACGCTCATTTAAATCTCACAGGG 2316
      ::         |||    ||||||
 602 laLysAspAsnSerAspLeu............................. 608

2317 CTTGCCACACTCAACGGCAATCTTAGTGCAAATGGCGATACACGTTATAC 2366
                :::    |||:::    |||:::::::::::::::
 609 .........IleIleGlyLysAlaSerSerAspAsnSerAsnAlaLysGl 622

2367 AGTCAGCCACAACGCCACCCAAAACGGCAACCTTAGCCTCGTGGGCAATG 2416
       ::::::    :::    :::::::::::::::::|||    :::||||||
 622 nIleThrPheAspLysValLysAspSerLysIleSer...AlaGlyAsnH 638

2417 CCCAAGCAACATTTAATCAAGCCACATTAAACGGCAACACATCGGCTTCG 2466
       ::::::|||:::|||:::    :::    ::::::|||:::::::::|||
 638 isAsnValThrLeuAsnSerLysValGluThrSerAsnSerAspGlySer 654

2467 ...GGCAATGCTTCA.................................. 2478
        ||||||:::|||
 655 ThrGlyAsnGlySerAspAspAsnAsnIleGlyLeuThrIleSerAlaLy 671

2479 ................TTTAATCTAAGCGACCACGCCGTACAAAACGGCA 2512
                        |||:::::::::|||    :::    |||    |
 671 sAspValThrValAsnSerAsnIleThrSerHisLysThrValAsnIleS 688

2513 GTCTGACGCTTTCCGGC...AACGCTAAGGCAAACGTAAGCCATTCCGCA 2559
      ||    :::    :::|||    :::::|||||||::::::::    :::|||
 688 erAlaSerGluGlyGlyIleThrThrLysAlaGlyThrThrIleAsnAla 704

2560 CTCAACGGTAATGTCTCCCTAGCCGATAAGGCAGTATTCCATTTTGAAAG 2609
       :::|||:::|||::::::::    |||
 705 ThrThrGlySerValGluValThrAlaLys................... 714

2610 CAGCCGCTTTACCGGACAAATCAGCGGC.........GGCAAGGATACGG 2650
                ||||||:::||||||||           ||||||    ::::
 715 .........ThrGlyAspIleSerGlyThrIleSerGlyLysThrValS 728

2651 CATTACACTTAAAAGACAGCGAATGGACGCTGCCGTCAGGCACGGAATTA 2700
      ::::::          ::::::    |||:::    :::|||:::::::::
 728 erValThrAlaThrThrAspSerLeuThrValLysGlyGlyAlaLysIle 744

2701 GGCAATTTAAACCTTGACAACGCCACCATTACACTCAATTCCGCCTATCG 2750
       |||    :::       :::||||||:::|||    :::|||:::
 745 ...AsnAlaThrGluGlyThrAlaThrLeuThrAlaSerSerGly..... 758
```

```
2751 CCACGATGCGGCAGGGGCGCAAACCGGCAGTGCGACAGATGCGCCGCGCC 2800
                    :::::::::::||||||   ::::::   :::
 759 .......LysLeuThrThrGluAlaAsnSerAlaIleSerGlyAlaAsnG 773

2801 GCCGTTCGCGCCGTTCGCGCCGTTCCCTATTATCCGTTACA......... 2841
         :::       |||:::   :::   :::|||   |||
 773 lyValThrAlaSerSerGlnSerGlyAspIleSerGlyThrIleSerGly 789

2842 CCGCCAACTTCGGTAGAATCCCGTTTCAACACGCTGACGGTAAACGGC.. 2889
        :::||||||   :::       ::::::||||||||||:::|||
 790 LysThrValSerValThrAlaSerSerGlySerLeuThrValGlyGlyAs 806

2890 ....AAATTGAACGGT...CAGGGAACATTCCGCTTTATGTCGGAACTCT 2932
         |||:::|||:::   :::|||:::       :::   :::
 806 pAlaLysIleAsnAlaThrGluGlyAlaAlaThrLeuThrAlaThrLysG 823

2933 TCGGCTACCGCAGCGACAAATTGAAGCTGGCGGAAAGTTCCGAAGGCACT 2982
                 :::   |||   :::       :::::::::|||||||||
 823 lyThrLeuThrThrValLysGlySerAsnIleAspAlaAsnGluGlyThr 839

2983 TACACCTTG...GCGGTCAACAATACCGGCAACGAACCTGCAAGCCTCGA 3029
        ::::::   |||   :::   |||   |||      ||||||    ::
 840 LeuValIleAsnAlaGlnAspAlaThrLeuAsnGlyAspAlaSerGlyAs 856

3030 ACAATTGACGGTA............GTGGAAGGAAAAGACAAC......A 3061
      ::::      |||             ::::::|||:::   |||
 856 pArgThrGluValAsnAlaValAsnAlaSerGlySerGlyAsnValThrA 873

3062 AACCGCTGTCCGAAAACCTTAATTTCACCCTGCAAAACGAACACGTCGAT 3111
               |||:::::::::|||:::|||    :::    :::   ::::::
 873 laLysThrSerSerSerValAsnIleThrGlyAspLeuSerThrIleAsn 889

3112 GCCGGCGCGTGGCGTTACCAACTCATCCGCAAAGACGGCGAGTTCCGCCT 3161
     :::             ::::::|||   |||:::|||:::
 890 Gly............LeuAsnIleIleSerLysAsnGlyLys........ 899

3162 GCATAATCCGGTCAAAGAACAAGAGCTTTCCGACAAACTCGGCAAGGCAG 3211
         |||   |||                          |||      ||||
 900 ....AsnThrVal.....................ValLeuLysGlyAlaG 908

3212 AAGCCAAAAAACAGGCGGAAAAAGACAACGCGCAAAGCCTTGACGCGCTG 3261
      ||         :::      :::   ::::::   |||   :::   :::
 908 luIleAspValLysTyrIleGlnProGlyValAlaSerAlaAsnGluVal 924

3262 ATTGCGGCCGGGCGCGATGCCGTCGAAAAGACAGAAAGCGTTGCCGAACC 3311
      |||   |||   |||    |||:::||||||:::::::::::::::::
 925 IleGluAlaLysArg...AlaLeuGluLysValLysAspLeuSerAspGl 940

3312 GGCCCGGCAGGCAGGCGGGGAAAATGTCGGCATTATGCAGGCGGAGGAAG 3361
         |||:::::   ::::::    :::|||:::         :::
 940 uGluArgGluThrLeuAlaLys...LeuGlyValSerAlaVal....... 953

3362 AGAAAAAACGGGTGCAGGCGGATAAAGACACCGCCTTGGCGAAACAGCGC 3411
         :::   ::::::   ::::::       ::::::      |||:::
 954 .....ArgPheIleGluProAsnAsnThrIleThrValAsnThrGlnAsn 968

3412 GAAGCGGAAACCCGGCCGGCTACC 3435
      |||      |||||||||::::::
 969 GluPheThrThrArgProSerSer 976
```

seq_name: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:AAB01832

```
seq_documentation_block:
ID   AAB01832 standard; Protein; 1011 AA.
XX
AC   AAB01832;
XX
DT   11-SEP-2000  (first entry)
XX
DE   Haemophilus influenzae strain K21 HMW2A protein, SEQ ID NO:39.
XX
KW   HMW protein; hmw gene; hmwA1; hmwA2; high molecular weight;
KW   non-typeable Haemophilus influenzae; NTHi; non-encapsulated;
KW   recombinant production; Escherichia coli; antibacterial; vaccine;
KW   human disease; otitis media; epiglottitis; pneumonia; tracheobronchitis;
KW   detection; diagnosis.
XX
OS   Haemophilus influenzae strain K21.
XX
PN   WO200020609-A2.
XX
PD   13-APR-2000.
XX
PF   07-OCT-1999;   99WO-CA00938.
XX
PR   07-OCT-1998;   98US-0167568.
PR   08-DEC-1998;   98US-0206942.
XX
PA   (CONN-) CONNAUGHT LAB LTD.
XX
PI   Loosmore SM,  Yang Y,  Klein MH;
XX
DR   WPI; 2000-303789/26.
DR   N-PSDB; AAA52181.
XX
PT   Nucleic acid molecule for producing recombinant high molecular weight
PT   proteins of Haemophilus which are used as a vaccine to provide
PT   protection against Haemophilus induced diseases in humans -
XX
PS   Claim 12; Fig 21A-O; 307pp; English.
XX
CC   The invention relates to the recombinant production of Haemophilus
CC   influenzae high molecular weight (HMW) proteins in Escherichia coli. The
CC   expression construct used to effect recombinant expression comprises a
CC   promoter functional in E. coli (e.g., the T7 promoter) operably linked
CC   to a modified hmwABC operon from a non-typeable (non-encapsulated) H.
CC   influenzae (NTHi). Most HMW-expressing NTHi strains contain two hmw gene
CC   clusters termed hmw1ABC and hmw2ABC. Each hmwABC operon comprises hmwA,
CC   hmwB and hmwC genes. The hmwA genes encode the structural HMWA proteins
CC   and the hmwB and hmwC genes encode accessory proteins which are
CC   responsible for post-translational processing and secretion of the HMWA
CC   proteins. The modified hmwABC operon used in the expression construct of
CC   the invention contains an A gene modified such that it encodes only the
CC   mature HMWA. The invention also discloses hmwA genes (AAA52175-A52198)
CC   and HMWA proteins (AAB01824-B01849) from the non-typeable H. influenzae
CC   strains Joyc, K1, K21, LCDC2, PMH1, 15 and 12. The nucleic acids and
CC   vectors are used for the production of recombinant H. influenzae HMW
CC   proteins which can be used as vaccines to mediate a humoral or
CC   cell-mediated immune response to provide protection against diseases in
CC   humans caused by H. influenzae (e.g., otitis media, epiglottitis,
CC   pneumonia and tracheobronchitis). The HMW proteins are also useful as
CC   antigens in immunoassays for detecting antibodies against Haemophilus,
CC   HMW proteins and/or HMW peptides. The nucleotide sequences encoding the
CC   HMW proteins can be used to isolate and clone hmw genes from other
CC   non-typeable strains of Haemophilus via hybridisation reactions. The
CC   present sequence represents an HMWA protein from a non-typeable strain of
CC   H. influenzae.
XX
SQ   Sequence   1011 AA;

alignment_scores:
           Quality:  259.50              Length:   1159
             Ratio:  0.451                 Gaps:     54
 Percent Similarity:  49.698   Percent Identity: 20.276

alignment_block:
 US-09-303-518D-649 x AAB01832   ..

 Align seg 1/1  to: AAB01832  from: 1  to: 1011

    226 AAAAAAGGGGAGTTGGTCGGCAAATCAATGACAAAAGCCCCGATGATTGA 275
           :::||||||      |||:::   |||   ||||||         :::::
     51 GlnLysGlyIleGluValAsnIleSerAlaThrLysAsnValThrValAs 67
```

```
 276 TTTTCTGTGGTGTCGCGTAACGGCGTGGCGGCATTGGTGGGCGATCAAT 325
     :   :::|||      :::||||||:::    :::|||   :::::::::
  67 nAlaAspValAspValLysAsnGlyThrLeuValLeuHisSerGlnArgA 84

 326 ATATTGTGAGCGTGGCACATAACGGCGGCTATAACAACGTTGATTTTGGT 375
           |||:::::      ||||||:::  ||||||:::
  84 snGlyValLysIle......AsnGlyAsn.IleThrSerThrGln..... 96

 376 GCGGAAGGAAGAAATCCCGATCAACATCGTTTTACTTATAAAATTGTGAA 425
                                                      |
  97 .................................................A 97

 426 ACGGA...ATAATTATAAAGCAGGGACTAAAGGCCATCCTTATGGCGGCG 472
     |||||    :::   ||||||   |||   |||:::    :::
  97 snGlyAsnLeuThrIleLysThrGlyGlyLysValAspValHisLysAsn 113

 473 ATTATCATATGCCGCGTTTGCATAAATTTGTCACAGATGCAGAACCTGTT 522
     |||    :::               :::|||:::::::::   :::|||:::
 114 IleThrLeuGlyMetGlyPheLeuAsnIleThrSerAspAsnAsnIle.. 129

 523 GAAATGACCAGTTATATGGATGGGCGGAAATATATCGATCAAAATAATTA 572
                :::::::     |||    |||:::::::|||
 130 .........ThrPheGluLysGlyAspAsnLeuThrIleThr....... 140

 573 CCCTGACCGTGTTCGTATTGGGGCAGGCAGGCAATATTGGCGATCTGATG 622
                                 |||||||||     |||:::
 141 ......................AlaGlnGlyAsnIle..IleSerAsnG 149

 623 AAGATGAGCCCAATAACCGCGAAAGTTCATATCATATTGCAAGTGCGTAT 672
     ::::::::   :::    |||    |||:::       :::
 149 lnGluAsnLysGlnLeuArgPheSerAsnValSerLeu............ 161

 673 TCTTGGCTCGTTGGTGGCAATACCTTTGCACAAAATGGATCAGGT..... 717
     :::    :::    :::|||   |||||||:::   |||    ::::::
 162 AsnGlyMetGlyAlaGlyLeuThrPheThrAlaAsnLysGlyAsnHisTh 178

 718 ..........GGTGGCACAGTCAACTTAGGTAGTGAAAAAATT......A 751
                |||||||:::|||::::::::::::    :::      :
 178 rHisLysPheAspGlyThrLeuAsnIleSerGlyLysValValIleAsnG 195

 752 AACATAGCCCATATGGTTTTTACCAACAGGAGGCTCA............ 789
     ::    :::|||:::::::::    |||    ::::::|||
 195 lnThrThrProHisAsnIleAlaProTrpAsnAlaSerAlaAspSerTyr 211

 790 .....................TTTGGCGACAGTGGCTCACCAATGTTTAT 818
                         :::|||::::::::::::      |||||
 212 TrpAsnValThrThrLeuThrLeuGlyAsnAsnAlaGlnPheThrPheIl 228

 819 C......TATGATGCCCAAAAGCAAAAGTGGTTAATTAATGGGGTATTGC 862
     |            |||:::::::::::          |||    :::|||          :
 228 eLysPheValAspSerAsnArgSerValAlaLeuAsnSerGlySerArgS 245

 863 AA...ACGGGCAACCCCTATATAGGAAAAAGCAATGGCTTCCAGCTGGTT 909
     ::    :::|||         :::    ||||||:::|||    :::
 245 erPheAlaGlyValLysPheTyrGlyLysAsnAsnGluMet......... 258

 910 CGTAAAGATTGGTTCTATGATGAAATCTTTGCTGGAGATACCCATTCAGT 959
                             ::::::       |||||||:::    :::||
 259 .....................LysPheAsnIleGlyAspAsnAlaAsnVa 268

 960 ATTCTACGAACCACGTCAAAATGGGAAATACTCTTTTAACGAC....... 1002
     |    ::::::    ::::::|||    :::    |||    |||
 268 lGluPheLysLeuLysSerAsnAspAsnThrSerAsnAsnLysProLeuP 285

1003 ............................GATAATAATGGCACAGGAAAA 1023
                                 :::|||||||||     :::
 285 roIleGlnPheLeuSerAsnIleSerAlaThrGlyAsnGlyThrValSer 301

1024 ATCAATGCCAAACATGAACACAATTCTCTGCCTAATAGATTAAAAACACG 1073
     ::::::              |||    :::|||        |||
 302 PheAspIle.........HisAlaAsnLeuSerAlaArgSerThrGluLe 315

1074 AACCGTTCAATTGTTTAATGTTTCT......................... 1098
     ::::::::|||:::|||:::|||
 315 uAsnMetSerLeuIleAsnIleSerAsnGlyValAsnPheSerIleAsnS 332

1099 ....................TTATCCGAGACAGCAAGAGAACCTGTTTAT 1128
                          |||      ::::::     :::
 332 erHisValArgGlyAsnAsnAlaPheGluIleLysLysAspLeuIleIle 348

1129 CATGCTGCAGGTGGTGTCAACAGTTATCGACCCAGACTGAATAATGGAGA 1178
     :::|||:::|||:::          :::    :::          ::::::     ::
 349 AsnAlaThrGlySerAsnPheAsnLeuLysGlnThrLysAspLysPheAs 365

1179 AAATATTTCCTTTATTGACGAAGGAAAAGGCGAATTGATACTTACCAGCA 1228
     :|||              :::       |||:::    ::::::    |||     |
 365 pAsn............SerTyrGluLysAsnAlaIlePheSerThrHisA 378

1229 ACATCAATCAAGGTGCTGGAGGATTATATTTCCAAGGAGATTTTACGGTC 1278
     ||::::::                        :::    |||:::    |||:::
 378 snLeuThr.................IleLeuGlyGlyAsnValThrLeu 388

1279 TCGCCTGAAAATAACGAAACTTGGCAAGGCGCGGGCGTTCATATCAGTGA 1328
     :::    ||||||:::::::::          ::::::::::::|||::::::
 389 GlyGlyGluAsnSerSerSerAsnIleLysGlyAsnIleAsnIleAsnSe 405

1329 AGACAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAAACGACCGCCTGT 1378
     :    ::::::||||||    ::::::::::::
 405 rLysAlaAsnValThrLeuGlnAlaHisAla.................. 415

1379 CCAAAATCGGCAAAGGCACGCTGCACGTTCAAGCCAAAGGGGAA...... 1422
                   ||||||    |||:::::::    |||    :::
 416 .............GlyThrSerHisLeuAspLysLysGluArgThrLeu 427

1423 AACCAAGGCTCGATCAGCGTGGGCGACGGTACAGTCATTTTGGATCAGCA 1472
     :::    |||::::::||||||||||    :::          |||:::    :::::
 428 ThrLeuGlyAsnValSerValGlyGlyAsnLeuAsnIleIleGlySerAs 444

1473 GGCAGACGATAAAGGCAAAAAACAAGCCTTTAGTGAAATCGGCTTGGTCA 1522
     :|||         |||    :::    ::::::::|||    :::         :
 444 nAlaHisIleAspGly...AsnLeuSerIleAlaGluSerAlaLysPheG 460

1523 GCGGCAGGGGTACGGTGCAACTGAATGCCGATAATCAGTTCAACCCCGAC 1572
     ::|||:::    :::    :::||||||          :::    |||:::
 460 lnGlyLysThrAsnAsnAsnLeuAsnIleThrGlyThrPheThr...... 474

1573 AAACTCTATTTCGGCTTTCGCGGCGGACGTTTGGATTTAAACGGGCATTC 1622
                         ::::::|||       |||:::|||
 475 ..................AsnAsnGlyThrAlaAspIleAsn........ 482

1623 GCTTTCGTTCCACCGTATTCAAAATACCGATGAAGGGGCGATGATTGTCA 1672

 482 .................................................. 482

1673 ACCACAATCAAGACAAAGAATCCACCGTTACCATTACAGGCAATAAAGAT 1722
                    |||::::::::|||     :::     |||
 483 ..........IleLysGlnGlyValValLysLeuGlnGly......... 492

1723 ATTGCTACAACCGGCAATAACAACAGCTTGGATAGCAAAAAAGAAATTGC 1772

 492 .................................................. 492

1773 CTACAACGGTTGGTTTGGCGAGAAAGATACGACCAAAACGAACGGCGGCGC 1822
                             |||     |||     :::||||||||:::|
 493 ........................AspIleThr...AsnAsnGlyAsnL 500

1823 TCAACCTTGTTTACCAGCCCGCCGCAGAAGACCGCACCCTGCTGCTTTCC 1872
     |||||::::::    :::    :::::::::::::::|||:::
 500 euAsnIleThrThrAsnAlaSerValAsnGlnLysThrIle......... 513

1873 GGCGGAACAAATTTAAACGGCAACATCACGCAAACAAACGGCAAACTGTT 1922
                   :::|||||||||||||||:::      :::|||     |||
 514 ............IleAsnGlyAsnIleThrAsnLysLysGlyAspLeuAs 526

1923 TTTCAGCGGCAGACCAACACCGCACGCCTACAATCATTTAAACGACCATT 1972
     ::::::          :::     :::|||          ::::::::     :::
```

```
 526 nIleLysAspIleLysAla...AsnAlaGluIleGlnIleGlyGlyAsnI 542

1973 GGTCGCAAAAAGAGGGCATTCCTCGCGGGGAAATCGTGTGGGACAACGAC 2022
       |||||||||||              |||:::::::::   ::::::|||
 542 leSerGlnLysGlu............GlyAsnLeuThrIleSerSerAsp 554

2023 TGGATCAACCGCACATTTAAAGCGGAAAACTTCCAAATTAAAGGCGGACA 2072
        ||||||   |||   :::   |||
 555 LysIleAsnIleThrLysArgIleGlu...................... 563

2073 GGCGGTGGTTTCCCGCAATGTTGCCAAAGTGAAAGGCGATTGGCATTTGA 2122
                                 :::|||:::|||        :
 564 ...........................IleLysAlaAsp.........T 568

2123 GCAATCACGCCCAAGCAGTTTTGGTGTCGCACCGCATCAAAGCCACACA 2172
     ::::::::::::::::        |||||||||   :::   :::   |||
 568 hrAspGlnGlyAsnSerAspSerGlyValAlaSerAsnAlaAsnLeuThr 584

2173 ATCTGTACACGTTCGGACTGGACGGGTCTGACAAATTGTGTCGAAAAAAC 2222
     |||   |||:::                               |||   ||
 585 IleLysThrLys.............................GluLeuTh 591

2223 CATTACCGACGATAAAGTGATTGCTTCATTGACTAAGACCGACATCAGCG 2272
     |:::||||||:::      |||:::::::::::::|||::::::|||::::
 591 rLeuThrAspAsnLeuAsnIleSerGlyPheAsnLysAlaGluIleThrA 608

2273 GC......AATGTCGATCTTGCCGATCACGCTCATTTAAATCTCACAGGG 2316
     ::          |||   ||||||
 608 laLysAspAsnSerAspLeu............................. 614

2317 CTTGCCACACTCAACGGCAATCTTAGTGCAAATGGCGATACACGTTATAC 2366
              :::   |||:::   |||::::::::::::::::
 615 .........IleIleGlyLysAlaSerSerAspAsnSerAsnAlaLysGl 628

2367 AGTCAGCCACAACGCCACCCAAAACGGCAACCTTAGCCTCGTGGGCAATG 2416
      ::::::   :::   ::::::::::::::::::|||   :::||||||
 628 nIleThrPheAspLysValLysAspSerLysIleSer...AlaGlyAsnH 644

2417 CCCAAGCAACATTTAATCAAGCCACATTAAACGGCAACACATCGGCTTCG 2466
       ::::::|||:::|||:::   :::   ::::::|||:::::::::|||
 644 isAsnValThrLeuAsnSerLysValGluThrSerAsnSerAspGlySer 660

2467 ...GGCAATGCTTCA.................................. 2478
        ||||||:::|||
 661 ThrGlyAsnGlySerAspAspAsnAsnIleGlyLeuThrIleSerAlaLy 677

2479 |...............TTTAATCTAAGCGACCACGCCGTACAAAACGGCA 2512
                      |||::::::::::|||   :::   |||   |
 677 sAspValThrValAsnSerAsnIleThrSerHisLysThrValAsnIleS 694

2513 GTCTGACGCTTTCCGGC...AACGCTAAGGCAAACGTAAGCCATTCCGCA 2559
     ||   :::   :::|||   ::::::||||||:::::::::   :::|||
 694 erAlaSerGluGlyGlyIleThrThrLysAlaGlyThrThrIleAsnAla 710

2560 CTCAACGGTAATGTCTCCCTAGCCGATAAGGCAGTATTCCATTTTGAAAG 2609
        :::|||:::|||::::::::   |||
 711 ThrThrGlySerValGluValThrAlaLys.................... 720

2610 CAGCCGCTTTACCGGACAAATCAGCGGC.........GGCAAGGATACGG 2650
               ||||||:::|||||||||          ||||||   ::::
 721 ..........ThrGlyAspIleSerGlyThrIleSerGlyLysThrValS 734

2651 CATTACACTTAAAAGACAGCGAATGGACGCTGCCGTCAGGCACGGAATTA 2700
     :::::           ::::::   |||:::   :::|||:::::::::
 734 erValThrAlaThrThrAspSerLeuThrValLysGlyGlyAlaLysIle 750

2701 GGCAATTTAAACCTTGACAACGCCACCATTACACTCAATTCCGCCTATCG 2750
        |||   :::       :::|||||||:::|||   :::|||:::
 751 ...AsnAlaThrGluGlyThrAlaThrLeuThrAlaSerSerGly..... 764

2751 CCACGATGCGGCAGGGGCGCAAACCGGCAGTGCGACAGATGCGCCGCGCC 2800
                 ::::::::::::||||||   ::::::   :::
 765 .......LysLeuThrThrGluAlaAsnSerAlaIleSerGlyAlaAsnG 779

2801 GCCGTTCGCGCCGTTCGCGCCGTTCCCTATTATCCGTTACA......... 2841
        :::      |||:::   :::   :::|||   |||
 779 lyValThrAlaSerSerGlnSerGlyAspIleSerGlyThrIleSerGly 795

2842 CCGCCAACTTCGGTAGAATCCCGTTTCAACACGCTGACGGTAAACGGC.. 2889
          :::||||||   :::      ::::::||||||||||:::|||
 796 LysThrValSerValThrAlaSerSerGlySerLeuThrValGlyGlyAs 812

2890 ....AAATTGAACGGT...CAGGGAACATTCCGCTTTATGTCGGAACTCT 2932
        |||:::|||:::   :::|||:::      :::   :::
 812 pAlaLysIleAsnAlaThrGluGlyAlaAlaThrLeuThrAlaThrLysG 829

2933 TCGGCTACCGCAGCGACAAATTGAAGCTGGCGGAAAGTTCCGAAGGCACT 2982
               :::   |||   :::      :::::::::|||||||||
 829 lyThrLeuThrThrValLysGlySerAsnIleAspAlaAsnGluGlyThr 845

2983 TACACCTTG...GCGGTCAACAATACCGGCAACGAACCTGCAAGCCTCGA 3029
        ::::::   |||   :::   |||   |||      ||||||   ::
 846 LeuValIleAsnAlaGlnAspAlaThrLeuAsnGlyAspAlaSerGlyAs 862

3030 ACAATTGACGGTA............GTGGAAGGAAAAGACAAC......A 3061
     ::::      |||            ::::::|||:::   |||
 862 pArgThrGluValAsnAlaValAsnAlaSerGlySerGlyAsnValThrA 879

3062 AACCGCTGTCCGAAAACCTTAATTTCACCCTGCAAAACGAACACGTCGAT 3111
              |||:::::::::|||:::|||   :::   :::   ::::::
 879 laLysThrSerSerSerValAsnIleThrGlyAspLeuSerThrIleAsn 895

3112 GCCGGCGCGTGGCGTTACCAACTCATCCGCAAAGACGGCGAGTTCCGCCT 3161
     :::               ::::::|||   |||:::|||:::
 896 Gly............LeuAsnIleIleSerLysAsnGlyLys........ 905

3162 GCATAATCCGGTCAAAGAACAAGAGCTTTCCGACAAACTCGGCAAGGCAG 3211
         |||   |||                     |||      ||||
 906 ....AsnThrVal....................ValLeuLysGlyAlaG 914

3212 AAGCCAAAAAACAGGCGGAAAAAGACAACGCGCAAAGCCTTGACGCGCTG 3261
     ||         :::      :::   ::::::   |||   :::   :::
 914 luIleAspValLysTyrIleGlnProGlyValAlaSerAlaAsnGluVal 930

3262 ATTGCGGCCGGGCGCGATGCCGTCGAAAAGACAGAAAGCGTTGCCGAACC 3311
     |||   |||   |||   |||:::||||||:::::::::::::::::::
 931 IleGluAlaLysArg...AlaLeuGluLysValLysAspLeuSerAspGl 946

3312 GGCCCGGCAGGCAGGCGGGGAAAATGTCGGCATTATGCAGGCGGAGGAAG 3361
        |||::::::   ::::::   :::|||:::         :::
 946 uGluArgGluThrLeuAlaLys...LeuGlyValSerAlaVal....... 959

3362 AGAAAAAACGGGTGCAGGCGGATAAAGACACCGCCTTGGCGAAACAGCGC 3411
          :::   ::::::   ::::::      ::::::      |||:::
 960 .....ArgPheIleGluProAsnAsnThrIleThrValAsnThrGlnAsn 974

3412 GAAGCGGAAACCCGGCCGGCTACC 3435
     |||      |||||||||::::::
 975 GluPheThrThrArgProSerSer 982
```